

# Forecasting the South African Rand's variance and covariance using Conditional Heteroskedastic and Realized Volatility models.

Christopher Sumter\*

UCT School of Economics

June 27, 2013

## Abstract

The last 30 years has seen the proliferation of numerous models that incorporate conditional heteroskedasticity. These models have been used extensively within financial economics to explain and forecast the volatility of financial instruments. However, recently, the realized volatility literature has yielded arbitrarily precise estimators of the ex-post volatility by exploiting high-frequency data. It has been shown that when allied with traditional time-series techniques such as ARFIMA models, these improved volatility estimators are able to provide superior volatility forecasts to that of the well-established GARCH methodology. We specify several conditional heteroskedasticity models for the USDZAR exchange rate and compare their out-of-sample forecasting results to those of the realized volatility model, in both a univariate and multivariate setting. In both instances we make use of "robust" statistical loss functions whose ranking of the respective models is invariant to the usage of a volatility proxy. Within the multivariate setting, we also introduce a new method for evaluating the output of these models by ranking them based on their ability to minimize risk. The results suggest that the univariate forecasts from the realized volatility model are unequivocally superior to a number of GARCH specifications. In the multivariate setting the evidence is slightly less convincing. This is particularly apparent in the case of the economic loss function where traditional MGARCH models are able to provide a comparable degree of risk reduction to that of the realized volatility form model.

## 1 Introduction

Correct modelling of the variance of a financial asset is crucial as it intimately relates to that asset's risk. The Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized ARCH (GARCH) literature beginning

---

\*I would like to thank my supervisor Kevin Kotze for his invaluable contribution and assistance, as well as my colleague Valeri Sokolovski, without whom this thesis would never have been written. I would also like to thank my father, Leon Sumter, for his diligent editorial efforts.

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

with Engle (1982), lead to the development of a vast array of models of varying complexity that provide an explicit framework for describing the dynamic behavior of the conditional second-order moment of financial asset returns. However, Andersen and Bollerslev (1998) note that the early out-of-sample performance of the GARCH model of Bollerslev (1986) was poor when compared to its in-sample performance. They proposed a new measure for the ex-post volatility that should serve as the benchmark against which forecasts from any model are judged. By summing intraday returns constructed from high-frequency data they were able to show that a far more accurate measure of the latent volatility was obtained, and this brought about a concomitant improvement in the out-of-sample forecasting performance of the GARCH models. This "realized volatility" estimator, formalized in Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002) relies on assumptions that are violated in reality, chief amongst them the assumption that returns are i.i.d. Market microstructure effects, such as the bid-ask spread, introduce negative autocorrelation into the returns process, particularly at the higher frequencies that are used in the realized volatility<sup>1</sup> estimator, Hansen and Lunde (2006).

The continued development of volatility estimators that can deal with market microstructure noise as well as other empirical regularities in high-frequency data has yielded increasingly more accurate measures of ex-post volatility<sup>2</sup>. Concurrently, another strand of research has built upon the insight that if these ever-improving volatility estimators are able to provide arbitrarily precise measures of ex-post volatility, then perhaps the direct modelling of these estimates for the latent volatility will yield better forecasts for the variance or covariance. A series of highly influential papers by Torben Andersen and Tim Bollerslev along with various authors have characterized both the unconditional and conditional moments of these realized volatility estimates, and have provided interesting results on their forecasting abilities, relative to the traditional GARCH models<sup>3</sup>.

A further development has been the amalgamation of realized volatility estimators and the GARCH methodology whereby the basic structure of the GARCH model is maintained, but augmented with a measure of realized volatility as in Engle (2002b) and Engle and Gallo (2006). Alternatively the usual measure of the most recent period's volatility, the squared return, is replaced by the realized volatility, as in the HEAVY model of Shephard and Sheppard (2010) and the Realized GARCH of Hansen et al. (2012).

In terms of testing for differences in forecasting performance, standard statistical loss functions such as

---

<sup>1</sup>Throughout this paper we will use "realized volatility" as a generic term to describe either the realized variance or realized standard deviation calculated using any possible estimator. When necessary we will distinguish between the realized variance and realized standard deviation as well as different estimators.

<sup>2</sup>Kernel-based estimators or "realized kernels" account for the microstructure induced autocorrelation by estimating the sample autocorrelation, see Zhou (1996) , Hansen and Lunde (2006), Bandi and Russell (2011) and Barndorff-Nielsen et al. (2008). Range-based or Extreme Value estimators are also robust to microstructure noise, but do not provide the level of accuracy afforded by realized kernels, see Parkinson (1980), Garman and Klass (1980), Yang and Zhang (2000) and Alizadeh et al. (2002). Additional means of dealing with microstructure noise are the "pre-whitening" techniques of Andersen et al. (2001) , Corsi et al. (2003), and Hansen et al. (2008) that filter the raw returns data to get rid of the autocorrelation.

<sup>3</sup>See Andersen et al. (2001) , Andersen et al. (2001), Andersen et al. (2003) and Andersen et al. (2007).

the Mean Square Error (MSE) and Mean Absolute Deviation (MAD) have often been used in combination with tests for Equal Predictive Ability (EPA) such as the Diebold-Mariano (DM) test (Diebold and Mariano (1994)), or tests of Superior Predictive Ability (SPA) such as those discussed in White (2003) and Hansen (2005). However, as noted in Andersen et al. (2005) the usage of a proxy for the latent volatility necessitates that a loss function be robust in the sense that the ranking of a set of competing models is not affected by the usage of such a proxy. Accordingly, Patton (2011) and Laurent et al. (2012) separately derive a class of robust loss functions for univariate and multivariate applications respectively.

After reviewing the relevant literature on volatility models and the implications of long-memory processes, we then examine the unconditional and conditional distribution of daily USDZAR variance which is calculated using the Moving Average (MA)-based estimator of Hansen et al. (2008) and intraday 5-minute prices. Our findings corroborate those of Bollerslev and Ole Mikkelsen (1996) and Andersen et al. (2001) as the daily USDZAR variance is shown to follow a long-memory process. Using the Box-Jenkins methodology of Box and Jenkins (1970) and a general-to-specific approach we fit an appropriate Autoregressive Fractionally Integrated Moving Average (ARFIMA) model to the USDZAR variance. A series of 1-step-ahead forecasts of the USDZAR volatility from the ARFIMA model are then compared to a number of popular GARCH specifications.

The mix of GARCH models was chosen in such a way that key features of the ARFIMA model were also captured by these rival models. The FIGARCH model of Baillie et al. (1996) incorporates long-memory while the HEAVY model of Shephard and Sheppard (2010) makes use of a high-quality proxy for recent volatility. For completeness a traditional GARCH model is also fitted, however, an asymmetric volatility response parameter is added to this particular model as USDZAR volatility appears to be higher following a depreciation against the dollar as opposed to an appreciation.

In reality investors care less about the risk of a single asset and more about the risk of their entire portfolio. It is thus natural to examine the forecasting powers of a multivariate model from the point of view of an investor optimizing his existing portfolio with respect to her exposure to the currency risk. In the second part of the paper we extend our analysis into a bivariate setting by adding a series of historical returns from a local investor's hypothetical portfolio. The properties of the monthly USDZAR 1-month forward variance and covariance with the aforementioned portfolio are analyzed and modelled using an ARFIMA process and compared against a set of common Multivariate GARCH models.

Practical difficulties such as the "curse-of-dimensionality" and the need to ensure positive-definiteness complicate the process of modelling the covariance matrix. The MGARCH models selected in this paper represent a cross-section of all the major approaches to overcoming these obstacles. The BEKK model of Engle and Kroner (1995) is a generalization of the GARCH model that is defined in quadratic forms. The DCC model of Engle (2002a) parameterizes the correlation between assets directly and is flexible with regards to the process that each asset's volatility follows. The GOGARCH model of Van der Weide (2002) relies on an orthogonal transformation of unobserved factors that drive returns and is particularly useful when the

portfolio is large. Finally, the RiskMetrics model of Metrics (1997) exponentially smooths each element of the covariance matrix using a common parameter.

In the univariate and multivariate setting we only use robust loss functions to make our forecast comparisons. Furthermore in the spirit of West et al. (1993) and Fleming et al. (2003) we attempt to measure the economic gains a risk-minimizing investor might enjoy should he/she decide to hedge his currency exposure using forecasts of the conditional covariance matrix. Individually or collectively, the rival models are not able to better the forecasting performance of the ARFIMA model based on statistical measures. In the multivariate evaluation the case for the superiority of the ARFIMA model over the MGARCH models is slightly less compelling, particularly when the economic loss function is used to judge performance.

This paper makes three main contributions. Firstly, we conduct an empirical examination of USDZAR volatility at the daily and monthly frequency. Secondly, we provide an answer to the following question: What is the best means of forecasting USDZAR volatility? Finally, we introduce another means of comparing covariance forecasts by way of a currency-hedging local investor. The prior literature has focused on currency-hedging for an investor with foreign currency denominated assets. Here we show that one can rely on a contemporaneous relationship between the first moment of *domestic* asset prices and the first moment of the exchange rate in order to motivate the need for hedging against currency risk.

The paper proceeds as follows; Section Two reviews the Univariate and Multivariate GARCH literature, Section Three provides a brief review of the Realized Volatility literature and discusses the implications for volatility forecasting as well as forecast evaluation, Section Four discusses the forecast comparison methodology that will be utilized and introduces the currency-hedging framework. Section Five discusses the Data as well as the Results, Section Six concludes.

## 2 ARCH modelling

To properly place the development of ARCH modeling within the financial econometrics literature one has to take cognizance of some of the stylized facts for daily financial asset returns. Firstly, the distribution of successive price changes or returns is not normal. Secondly, there is mostly no statistically significant autocorrelation between returns. Thirdly, there is evidence of positive dependence between absolute returns or squared returns on nearby days.

In terms of the distribution of financial returns, the most common finding is that of leptokurtosis, or fat tails, in that extreme returns occur far more frequently than implied by the Gaussian distribution (Mandelbrot (1963), Fama (1965)). The positive dependence between squared returns and absolute returns which are both proxies for volatility can be explained by the volatility clustering or time-varying volatility first articulated in Mandelbrot (1963). By specifying some sort of a process for volatility these stylized facts are better understood. The vast family of ARCH models that began with the seminal work of Engle (1982) essentially apply traditional time series tools, such as the ARMA framework of Box and Jenkins (1970),

to the conditional variance instead of the mean (returns) of a process. What follows below is a review of univariate ARCH modelling with special emphasis placed on the models that will be estimated in this paper. A more thorough discussion can be found in Bollerslev et al. (1992), Bollerslev et al. (1994) and Engle (2001); although more recent developments such as FIGARCH, BEKK, DCC and HEAVY are not covered in those reviews.

Prior to the introduction of ARCH/GARCH intertemporal dependence in the second moment of returns would often be modelled by a rolling regression whereby the conditional standard deviation was an equally weighted average of the prior  $q$  standard deviations. It seems logical however that recent observations should be weighted more heavily. The ARCH framework goes one step further and allows the weights to be estimated and their significance to be tested. Given the information available at time  $t - 1$ ,  $I_{t-1}$  where this is typically the history of returns up until time  $t - 1$ , a general ARCH framework in discrete time<sup>4</sup> takes the following form;

Returns  $\{r_t\}_{t=1}^T$  are given by the change in the natural logarithm of prices  $\{p_t\}_{t=1}^{T+1}$  :

$$r_t = \mu_t + e_t = p_t - p_{t-1}. \quad (1)$$

The residuals  $\{e_t\}_{t=1}^T$  follow a discrete-time stochastic process. They are a function of the conditional variance  $\{h_t\}_{t=1}^T$ , the standardized residuals  $\{z_t\}_{t=1}^T$  and are assumed to be conditionally Gaussian:

$$e_t = z_t h_t^{1/2}, \quad (2)$$

where the  $\{z_t\}$  have zero mean and unit variance and are i.i.d.,  $z_t|I_{t-1} \sim n.i.d(0, 1)$  and thus  $e_t|I_{t-1} \sim N(0, h_t)$ . This in turn implies that returns are conditionally normally distributed,  $r_t|I_{t-1} \sim N(\mu_t, h_t)$  and that the standardized residuals  $\{z_t\}$  are given by:

$$z_t = \frac{r_t - \mu_t}{h_t^{1/2}}. \quad (3)$$

The mean,  $\mu_t$ , as well as the variance,  $h_t$ , are conditional on the information set at time  $t - 1$ , as well as the parameter vector,  $\theta$ , which will be populated shortly. Various processes for the conditional variance have been studied in the literature, commencing with the path-breaking paper of Engle (1982) where the conditional variance is a stochastic process and a linear function of the previous squared residuals<sup>5</sup>:

$$h_t = \omega + \alpha(e_{t-1})^2. \quad (4)$$

The autoregressive nomenclature for the conditional variance process is only apparent after some manipulation. Defining the forecast error when predicting squared residuals as:

---

<sup>4</sup>While the prices of assets traded in liquid financial markets evolve in a near-continuous fashion, most volatility models are formulated in discrete time and prices and returns are viewed as discrete observations from an underlying continuous-time process.

<sup>5</sup>Note that in all that follows, the mean  $\mu_t$  will be assumed to be zero.

$$\nu_t = e_t^2 - h_t, \quad (5)$$

and then substituting this into Equation (4) yields<sup>6</sup>:

$$e_t^2 = \omega + \alpha e_{t-1}^2 + \nu_t. \quad (6)$$

Thus the squared residuals follow an AR(1) process and are a linear function of the past sample variances  $e_{t-1}^2$ . Alternatively  $e_t^2$  could just as easily have followed an AR(q) or ARCH(q) process. Just as an AR process can be extended to an ARMA process, the same can be done for the ARCH model, which was generalized in Bollerslev (1986) in the form of the GARCH(p,q) model that supplements the lagged sample variances with  $p$  lagged conditional variances:

$$h_t = \omega + \sum_{i=1}^q \alpha_i e_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}. \quad (7)$$

Thus the parameter vector is  $\theta = [\alpha_1 \dots \alpha_q, \beta_1 \dots \beta_p]$ . To better understand the dynamics of the conditional variance consider the GARCH(1,1) model:

$$h_t = \omega + \alpha(e_{t-1})^2 + \beta(h_{t-1}). \quad (8)$$

After taking expectations of (8) and rearranging, the unconditional variance is given by:

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}. \quad (9)$$

Therefore under the GARCH specification the best predictor of  $h_t$  is a weighted average of the long run average variance  $\sigma^2$ , the variance predicted for the previous period  $h_{t-1}$  and the new information captured by the most recent squared residual  $e_{t-1}^2$ , where the weights are  $((1-\alpha-\beta), \beta, \alpha)$ <sup>7</sup>. An alternative representation is that the GARCH model utilizes a weighted average of past squared residuals but these weights never go to zero and the model has an ARCH( $\infty$ ) representation. Using lag operators and rearranging:

$$h_t = \frac{\omega}{1 - \beta} + \alpha \sum_{i=0}^{\infty} \beta^i e_{t-1-i}^2. \quad (10)$$

A necessary restriction to ensure  $h_t$  is always positive is that  $\alpha > 0, \beta > 0, \omega > 0$ . Bollerslev (1986) has shown that this process is covariance stationary if  $\alpha + \beta < 1$  or for the GARCH(p,q) if  $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$ . In terms of multi-step forecasts it can be shown that the conditional variance of the return  $n$  periods in the future can be calculated recursively as:

$$var(r_{t+n}|I_{t-1}) = \sigma^2 + (\alpha + \beta)^{n-1}(h_{t+1} - \sigma^2). \quad (11)$$

---

<sup>6</sup>Its worth noting at this point that one-step-ahead forecasts made from the GARCH model are optimal by construction in the mean square error sense as the expectation of the forecast error is zero.

<sup>7</sup>Engle (2001) likens this updating rule to adaptive or learning behavior.

The parameters  $\alpha$  and  $\beta$  determine how quickly the conditional variance forecasts revert back to the unconditional variance  $\sigma^2$  and their sum  $(\alpha + \beta)$  is termed the persistence parameter, which is denoted  $\phi$ . Thus given the stationarity of the process the conditional variance will revert geometrically back towards the unconditional variance as the forecast horizon  $n$  increases. Very often in empirical work the persistence parameter is estimated to be close to one. If however  $\alpha + \beta = 1$  then the model is of the Integrated-GARCH or IGARCH variety<sup>8</sup>, the properties of which are discussed in Engle and Bollerslev (1986)<sup>9</sup>:

$$h_t = \omega + \alpha(e_{t-1})^2 + (1 - \alpha)(h_{t-1}). \quad (12)$$

The IGARCH is not covariance stationary because it contains a unit root, and as such forecasts for the conditional variance  $n$  periods into the future are exactly the same as the  $n + 1$  forecast. Similar to a random walk with drift, shocks to the conditional variance have a permanent effect. A pure random walk model for the volatility however isn't backed up with empirical evidence with most studies suggesting only a near unit root process. Stationary GARCH processes will have an autocorrelation function (ACF) that displays exponential decay, but on the other hand the imposition of a unit root may exaggerate the long-run dependencies within the conditional variance process. A more realistic assumption and one that is backed up by empirical evidence would be that of a long memory process for the variance.

## 2.1 Long memory ARCH Models

As noted by Granger and Joyeux (1980), for any non-stationary Autoregressive Integrated Moving Average ARIMA(p,d,q) process it need not be the case that  $d$  is a positive integer. Intuitively it makes sense that the process for the conditional variance could have  $d < 1$ , as shocks may not have a permanent effect on the level of volatility. In their analysis of intraday FX volatility patterns, Dacorogna et al. (1993) found that the decline of the ACF for squared returns is slower than what would be implied by a stationary GARCH model. They further suggested that the empirical ACF showed hyperbolic decay as opposed to the exponential decay of its theoretical counterpart. Ding et al. (1993) have also shown that the same holds true for the S&P 500 equity index. If the ACF declines more slowly, then the process is said to be fractionally integrated and has long memory, in that distant shocks still have some effect on the process for the variance. Consider an ARMA(p,q) process for the returns  $\{r_t\}$  with zero mean and innovation  $\{e_t\}$ :

$$\phi(L)r_t = \theta(L)e_t, \quad (13)$$

where  $\phi(L)$  and  $\theta(L)$  are the lag polynomials<sup>10</sup> of order  $p$  and  $q$  respectively. Then an Autoregressive Fractionally Integrated Moving Average or ARFIMA(p,d,q) stationary process for the returns can be given by:

<sup>8</sup>For the GARCH(p,q) to be integrated  $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j = 1$

<sup>9</sup>The IGARCH model also has the added advantage of increasing the degrees of freedom.

<sup>10</sup> $\phi(L) = 1 - \sum_{i=1}^p \phi_i L^i$  and  $\theta(L) = 1 + \sum_{i=1}^q \theta_i L^i$



$$(1 - L)^d \phi(L) r_t = \theta(L) e_t, \quad (14)$$

where  $(1 - L)^d$  is defined by the infinite series expansion  $1 - dL + \frac{1}{2}d(d-1)L^2 - \dots$ . The process is fractionally integrated if  $d = [-0.5, 0.5]$ . Baillie et al. (1996) introduce the fractionally integrated GARCH or FIGARCH process for the residuals,  $e_t$ . Recalling the innovations to the conditional variance process (Equation (5)) then it can be shown that a FIGARCH(p,d,q) model has the following representation:

$$(1 - L)^d \alpha(L) e_t^2 = \omega + [1 - \beta(L)] \nu_t,$$

where  $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \dots$  and  $\beta(L) = \beta_1 L + \beta_2 L^2$  and with some relatively straightforward manipulation, the conditional variance of  $e_t$  is given by the infinite ARCH presentation:

$$\begin{aligned} h_t &= \frac{\omega}{1 - \beta_1} + [1 - (1 - \beta(L))^{-1} \alpha(L) (1 - L)^d] e_t^2, \\ &= \frac{\omega}{1 - \beta_1} + \lambda(L) e_t^2, \end{aligned} \quad (15)$$

where  $\lambda(L) = \lambda_1 L + \lambda_2 L^2 + \dots$ , and  $\lambda_i$  is a function of  $\lambda_{i-1}, \beta, \alpha$  and  $d^{11,12}$ .

## 2.2 Conditional Mean

Thus far we have only discussed the process for the conditional variance and not the mean. In contrast with the numerous parameterizations for the conditional variance, the number of processes for the conditional mean are relatively sparse and this is a direct result of stylized fact number two; the lack of correlation between daily returns. One of the fundamental tenets of financial theory is the relationship between risk and returns, all prominent theories in asset pricing such as the Capital Asset Pricing Model (CAPM) of Sharpe 1964 and the Arbitrage Pricing Theory (APT) of Ross (1973) have this relationship at their core. One such model that couples this principle with the ARCH methodology is the ARCH-in-mean model of Engle et al. (1987) which proposes that the conditional mean or expected return is a function of the conditional variance and this is consistent with markets where agents are risk averse and require compensation in the form of a risk premium for holding risky or volatile assets instead of a risk-free asset or the market portfolio depending on what the less risky benchmark may be<sup>13</sup>. This specification may not be entirely congruent with the currency

<sup>11</sup>A closely linked relative to the FIGARCH model is the Fractionally Integrated Exponential GARCH or FIEGARCH also to be found in Bollerslev and Ole Mikkelsen (1996) that builds on the EGARCH model. The main difference between the two models being that the FIEGARCH models the logarithm of  $h_t$  and the response to negative innovations is greater than that of positive innovations.

<sup>12</sup>Another approach to modelling long memory is given in Lee and Engle (1993) in which their Component GARCH model approximates long memory as the sum of a few individually short memory components.

<sup>13</sup>Engle et al. (1987) use the *ARCH-in-mean* model to show that investors require more compensation for holding 6 month treasury bills instead of 3 month t-bills in periods of increased volatility as measured by the conditional variance and that the relationship between expected returns and the conditional variance was statistically significant.

markets and here we consider only the simplest specification for the mean equation; that it is simply some constant. Examination of the data in Section Five will show this to be an accurate representation.

### 2.3 Asymmetric Innovations

Both the aforementioned ARCH and GARCH specifications treat positive and negative innovations symmetrically. In certain settings such as the equity markets negative shocks could have a greater impact on volatility than positive shocks as equity investors rush to liquidate their holdings in a falling market thereby manifesting greater volatility. Black (1976) was the first to report that stock returns and volatility are negatively correlated. Volatility tends to rise following returns lower than expected and fall following returns higher than expected. Nelson (1991) showed that a fall in the US stock market had a greater effect on future volatility than a rise of the same magnitude. In equity markets one theoretical explanation is the leverage effect whereby falling equity values raise the debt/equity ratio of the firm, increasing the risk level which then manifests itself as higher volatility, (Black (1976)). The empirical evidence to support the leverage effect is somewhat mixed where early evidence is reported in Christie (1982) and Duffee (1995). Another possible contributor to this phenomenon is the ‘volatility feedback’ effect which is also predicated on the concept of time-varying risk premia, French et al. (1987). Investors expect to be compensated with higher returns following increased volatility and this necessitates an immediate drop in prices. The Exponential GARCH (EGARCH) model of Nelson (1991), the GJR-GARCH model of Glosten et al. (1993) and the Threshold GARCH (TARCH) of Zakoian (1994) are three model specifications that incorporate the extra information embedded in the sign of the residual. The EGARCH model differs from the TARCH or GJR-GARCH in that the logarithm of the conditional variance is modelled instead and this ensures that large negative innovations have a larger impact on the conditional variance than positive innovations of the same size. TARCH and GJR-GARCH take a different route and utilize an indicator function  $I_{\{e_{t-o} < 0\}}$  that is 1 if  $e_{t-o} < 0$  and zero otherwise where  $o$  is the order of lagged asymmetric innovation terms. The expression below nests both the TARCH and GJR-GARCH as well as a number of other GARCH models<sup>14</sup>:

$$\sqrt{h_t}^\delta = \omega + \sum_{i=1}^q \alpha_i |e_{t-i}|^\delta + \sum_{k=1}^o \gamma_k |e_{t-o}|^\delta I_{\{e_{t-o} < 0\}} + \sum_{j=1}^p \beta_j \sqrt{h_{t-j}}^\delta. \quad (16)$$

Thus  $\gamma$  captures the effect of negative lagged innovations on  $h_t$ . TARCH models the conditional standard deviation  $\sqrt{h_t}$  (when  $\delta = 1$ ) while GJR-GARCH models the conditional variance  $h_t$  ( $\delta = 2$ ). The persistence parameter  $\phi$  is now given by  $(\alpha + \frac{1}{2}\gamma + \beta)$ <sup>15</sup>. There is no a priori reason for currency markets to display the same asymmetric volatility response that has been reported for the equity markets, as the leverage or volatility feedback effect are not congruent with these markets. However, an empirical examination of the

<sup>14</sup>Note that this is very similar to the Asymmetric Power ARCH (APARCH) of Ding et al. (1993), the only difference being that the order of asymmetric innovation terms need not be  $q$ .

<sup>15</sup>This can be deduced by writing  $h_t = \omega + ((\alpha + \gamma I)z_{t-1}^2 + \beta)h_{t-1}$  and taking expectations at time  $t - 1$ .

data suggests that the USDZAR rate does in fact respond differently to positive and negative returns, at least when one considers daily data.

## 2.4 Multivariate GARCH Models

Multivariate GARCH or MGARCH models allow one to investigate the extent to which the second moment of different assets are related to one another and may be used to test for volatility spillover effects or dynamic correlations between assets. Beyond these applications, MGARCH models have a number of practical applications in asset and derivative pricing, hedging and risk management. For example consider the asset pricing models that make use of factor-loadings/betas, or the sensitivity of asset  $i$  to market factor  $j$ . These betas are typically obtained by regressing assets  $i$ 's excess return<sup>16</sup> on the factor returns. If these factor-loadings are not constant through time then MGARCH models are best suited to estimate them since  $\beta$  is simply the covariance between the asset return and factor return divided by the variance of the factor return. Covariances and correlations also play a significant role in modern portfolio management. Forward-looking asset allocation and the determination of the optimal portfolio requires an accurate forecast of the covariance matrix which in turn can be used to calculate the correlation matrix. In terms of risk management, the Value-at-Risk (VaR) measures the probability of a loss exceeding a certain threshold for a portfolio of  $n$  assets. The VaR is crucially dependent on accurate measures of the conditional covariances and correlations. Ignoring for a moment the multivariate applications, MGARCH models can also potentially provide better forecasts of a particular asset's conditional variance by incorporating the conditional variance of another asset as well as the conditional covariance in the GARCH process.

Revisiting Equation (1) and now defining  $\mathbf{r}_t$ <sup>17</sup> as a  $n \times 1$  vector process where the mean vector  $\boldsymbol{\mu}$ <sup>18</sup> and the residual vector  $\mathbf{e}_t$  are of the same dimension, then returns are given by:

$$\mathbf{r}_t = \boldsymbol{\mu} + \mathbf{e}_t, \quad (17)$$

and the residual vector is given by:

$$\mathbf{e}_t = \mathbf{H}_t^{1/2} \mathbf{z}_t, \quad (18)$$

where  $\mathbf{H}_t^{1/2}$  is a  $n \times n$  positive definite matrix and is usually obtained through a Cholesky decomposition<sup>19</sup>. Each element of the vector  $\mathbf{z}_t$  is once again i.i.d. with mean zero and unit variance, i.e. the variance covariance matrix of  $\mathbf{z}_t$  is simply the identity matrix  $\mathbf{I}_n$ . This together implies that the conditional covariance matrix of  $\mathbf{r}_t$  is:

$$\text{var}(\mathbf{r}_t | I_{t-1}) = \mathbf{H}_t^{1/2} \text{var}(\mathbf{z}_t) (\mathbf{H}_t^{1/2})' = \mathbf{H}_t, \quad (19)$$

<sup>16</sup>The return less the risk free rate, usually a short term government security such as 3-month Treasury Bills.

<sup>17</sup>Here the bold typeface indicates a matrix or vector.

<sup>18</sup>Note that we have dropped the time subscript from  $\boldsymbol{\mu}$  given that its assumed to be constant.

<sup>19</sup>Loosely this can be thought of as the multivariate analogue to the square root operator.

where  $I_{t-1}$  is the price history of all  $n$  assets.

Much of the literature on MGARCH models has been theoretical in nature and focused mainly on two distinct issues; ways of ensuring that  $\mathbf{H}_t$  is positive definite for all  $t$  and ways of dealing with the "curse-of-dimensionality" whereby the number of parameters to be estimated becomes unmanageable for applications in which  $n$  is large. For most models positive-definiteness is enforced by a specific model structure allied with some constraints on the parameter vector.

Bauwens et al. (2006) classify existing MGARCH models into three non-mutually exclusive categories:

1. Direct generalizations of the univariate GARCH model of Bollerslev (1986).
2. Linear combinations of univariate GARCH models.
3. Nonlinear combinations of univariate GARCH models.

The following section reviews models from each of these categories with particular attention placed on models that will be estimated in later sections. A broader and more detailed exposition of the MGARCH literature can be found in Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2009).

#### 2.4.1 Direct Generalizations of the univariate GARCH model

**VEC** Bollerslev et al. (1988) propose the VEC(p,q) model with  $p$  lagged conditional variance terms and  $q$  lagged symmetric innovation terms. The VEC(1,1) model takes the following form:

$$\mathbf{h}_t = \mathbf{c} + \mathbf{A}\boldsymbol{\eta}_{t-1} + \mathbf{B}\mathbf{h}_{t-1}, \quad (20)$$

$$\mathbf{h}_{t-1} = \text{vech}(\mathbf{H}_{t-1}), \quad (21)$$

$$\boldsymbol{\eta}_{t-1} = \text{vech}(\mathbf{e}_{t-1}\mathbf{e}_{t-1}'), \quad (22)$$

$$\mathbf{c} = \text{vech}(\mathbf{C}), \quad (23)$$

where the *vech* operator simply stacks the lower triangular portion of an  $n \times n$  matrix as a  $\frac{n(n+1)}{2} \times 1$  vector.  $\mathbf{A}$  and  $\mathbf{B}$  are square parameter matrices of order  $n(n+1)/2$  and  $\mathbf{c}$  is a  $\frac{n(n+1)}{2} \times 1$  parameter vector<sup>20</sup>. Combining (20), (21), (22) and (23) yields the VEC model's matrix form:

$$\begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} c_{01} \\ c_{02} \\ c_{03} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} e_{1,t-1}^2 \\ e_{1,t-1}e_{2,t-1} \\ e_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix},$$

<sup>20</sup>The model is covariance-stationary provided the eigenvalues of  $\mathbf{A} + \mathbf{B}$  are less than one in absolute value while the unconditional covariance matrix is given by  $\boldsymbol{\Sigma} = [\mathbf{I}_{n(n+1)/2} - \mathbf{A} - \mathbf{B}]^{-1}\mathbf{C}$ .

and the conditional covariance for asset 1, is given by:

$$h_{11,t} = c_{01} + a_{11}e_{1,t-1}^2 + a_{12}e_{1,t-1}e_{2,t-1} + a_{13}e_{2,t-1}^2 + b_{11}h_{11,t-1} + b_{12}h_{12,t-1} + b_{13}h_{22,t-1}. \quad (24)$$

From the above expression we can clearly see that the VEC model does allow for spillover effects;  $h_{11,t}$  is a linear function of the lagged squared innovations, the lagged cross products of the innovations as well the lagged values of each assets conditional variance and the conditional covariance. The price of this rich specification is that for two assets, 21 parameters need to be estimated while for  $n$  assets  $n(n+1)(n+1)/2$  parameters must be estimated, thus moving from 2 to 3 assets requires a further 57 parameters to be estimated. The large number of parameters in VEC models along with the very restrictive conditions necessary to ensure that  $\mathbf{H}_t$  is positive definite has motivated simpler structures<sup>21</sup>. One such example is the the Diagonal VEC or DVEC model of Bollerslev et al. (1988) which is a special case of the VEC model. The  $\mathbf{A}$  and  $\mathbf{B}$  matrices are restricted to be diagonal, thereby greatly reducing the number of parameters to be estimated, specifically  $3/2n(n+1)$  parameters for  $n$  assets or 9 parameters when  $n = 2$ , as is illustrated below:

$$\begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} c_{01} \\ c_{02} \\ c_{03} \end{bmatrix} + \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} e_{1,t-1}^2 \\ e_{1,t-1}e_{2,t-1} \\ e_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix},$$

and  $h_{11,t}$  is reduced to a standard univariate GARCH expression:

$$h_{11,t} = c_{01} + a_{11}e_{1,t-1}^2 + b_{11}h_{11,t-1}. \quad (25)$$

The DVEC's relative parsimony is evident when comparing Equations (24) and (25), however by making the diagonal restriction, asset two's variance or the covariance doesn't appear in the conditional variance equation for asset 1, therefore some potentially material information to the volatility process of asset 1 may be omitted in the DVEC specification. One drawback of this particular model is the inability to ensure positive-definiteness of  $\mathbf{H}_t$ <sup>22</sup>.

**BEKK** The introduction of the BEKK<sup>23</sup> model of Engle and Kroner (1995) dealt with both of the VEC model's shortcomings by addressing the curse-of-dimensionality and providing a convenient means of tackling the positive-definiteness requirement. Their solution was to define the conditional covariance matrix in quadratic forms and make  $\mathbf{H}_t$  positive-definite by construction. Assuming  $p = q = 1$  the model is given by the following matrix expression:

---

<sup>21</sup>See Engle and Kroner (1995).

<sup>22</sup>The Flex-GARCH method of Ledoit et al. (2003) provides an interesting alternative to estimating a DVEC model whereby each possible combination of bivariate GARCH model is estimated and then transformed into parameter matrices in such a way as to ensure positive definiteness. However, for a larger number of assets this model too becomes cumbersome.

<sup>23</sup>The name is an acronym of the author's surnames from a prior paper: Baba, Engle, Kraft and Kroner

$$\mathbf{H}_t = \mathbf{C}'\mathbf{C} + \mathbf{A}'\mathbf{e}_{t-1}\mathbf{e}_{t-1}'\mathbf{A} + \mathbf{B}'\mathbf{H}_t\mathbf{B}, \quad (26)$$

where the matrices  $\mathbf{C}'\mathbf{C}$ ,  $\mathbf{A}$  and  $\mathbf{B}$  are  $n \times n$  and  $\mathbf{C}'\mathbf{C}$  is symmetric and positive definite, more specifically  $\mathbf{C}'\mathbf{C}$  is a Cholesky decomposition whereby the intercept in the model is decomposed into the product of two triangular matrices to ensure its positive-definiteness, as well as that of  $\mathbf{H}_t$ <sup>24</sup>. Much like the VEC model the BEKK model has its own restricted versions whereby Matrices  $\mathbf{A}$  and  $\mathbf{B}$  are diagonal or simply a scalar, see Ding and Engle (2001). The ‘Full’ version of the model allows for volatility spillovers and gives the following processes for the conditional variances:

$$h_{11,t} = c_{11}^2 + a_{11}^2 e_{1,t-1}^2 + 2a_{21}e_{1,t-1}e_{2,t-1} + a_{21}^2 e_{2,t-1}^2 + b_{11}^2 h_{11,t-1} + 2b_{11}b_{21}h_{12,t-1} + b_{21}^2 h_{22,t-1}. \quad (27)$$

Lagged residuals, conditional variances of the other assets as well as the conditional covariance between the two assets appear once again as in Equation (24), but the BEKK is more parsimonious because the parameters governing the dynamics of the covariance  $h_{12}$  are the products of the corresponding parameters of the variance equations for each of the two assets. If the model is reduced to its diagonal form in the same fashion as the DVEC the conditional variance for asset 1 doesn’t include any information from other assets and is equivalent to a standard univariate GARCH process once again.

**RiskMetrics** The sheer number of parameters more often than not hinders the practical application of the aforementioned MGARCH models. For example, portfolio managers typically deal with hundreds of assets which makes it nearly impossible to estimate all the parameters in a BEKK or DCC model. Practitioners are forced to fall back on simpler methods that are more feasible, one of these being the RiskMetrics methodology, (Metrics (1997)), which is based on exponential smoothing and is closely related to the IGARCH model. Andersen et al. (2002) note that the RiskMetrics model corresponds directly to a diagonal IGARCH(1,1) with all intercepts in  $\mathbf{H}_t$  equal to zero and where identical values for  $\alpha$  and  $\beta = 1 - \alpha$  are utilized for all  $n$  assets. Thus the RiskMetrics model prescribes the same random-walk or unit root process for each element of  $\mathbf{H}_t$ . Under this approach  $\alpha$  is known as the decay factor with recent observations given more weight than distant ones. In matrix notation the model can be described by:

$$\mathbf{H}_t = \alpha(\mathbf{e}_{t-1}\mathbf{e}_{t-1}') + (1 - \alpha)\mathbf{H}_t. \quad (28)$$

<sup>24</sup>The BEKK model is covariance stationary if the eigenvalues of  $\mathbf{A} \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{B}$  are less than one in absolute value, where  $\otimes$  defines the Kronecker product which is defined as:

$$\mathbf{A} \otimes \mathbf{A} = \begin{bmatrix} a_{11}\mathbf{A} & \cdots & a_{1n}\mathbf{A} \\ \vdots & \ddots & \vdots \\ a_{n1}\mathbf{A} & \cdots & a_{nn}\mathbf{A} \end{bmatrix}$$

If  $\mathbf{A}$  is  $n \times n$ .

The original paper proposes that the decay factor be 0.94 for daily data and 0.97 for monthly data. The drawbacks of this approach are obvious, as  $\alpha$  is somewhat arbitrarily selected and the dynamics for each conditional variance are extremely limited in that they all rely on the scalar  $\alpha$ . Unlike a Full BEKK or VEC model it does not allow for volatility spillovers. Regardless this approach is still widely used because of its lack of computational burden relative to other approaches. Positive-definiteness is also guaranteed provided the initial covariance  $H_o$  is the sample covariance and the outer product of residuals is positive semi-definite which is usually the case. The addition of the RiskMetrics model to our set of MGARCH models will provide an "industry-benchmark" of sorts when the out-of-sample forecasting abilities of all these models are assessed.

#### 2.4.2 Linear Combinations of Univariate GARCH Models

**Factor Models** The Factor-ARCH model of Engle et al. (1990) relies on economic theory instead of creative parameterizations to generate a more parsimonious model. F-ARCH rests on the Arbitrage Pricing Theory (APT) of Ross (1973). The data (returns) are generated by some unobserved components/factors (possibly correlated with one another) common to all assets, where the number of factors,  $k$ , is generally less than the number of assets,  $n$ . The F-ARCH(p,q,k) model assumes that these factors are conditionally heteroskedastic and follow a GARCH process. The degree to which the F-ARCH is easier to estimate relative to VEC or BEKK is entirely dependent on how many factors are needed to model returns. Suppose for the sake of illustration that there is only one factor,  $f_t$ , and the vector of returns,  $\mathbf{r}_t$ , is a linear function of this factor:

$$\mathbf{r}_t = \boldsymbol{\mu} + \boldsymbol{\beta}f_t + \mathbf{v}_t, \quad (29)$$

where  $\boldsymbol{\mu}$  is the  $n \times 1$  vector of intercepts and the error vector,  $\mathbf{v}_t$ , is assumed to be i.i.d with constant covariance matrix  $\boldsymbol{\Omega}$ . The asset-specific factor loadings are captured in the vector  $\boldsymbol{\beta}$ . If one defines the conditional variance of  $f_t$  as  $\sigma_t^2$ , then the conditional covariance for  $\mathbf{r}_t$  takes the following form:

$$\mathbf{H}_t = \boldsymbol{\Omega} + \boldsymbol{\beta}\boldsymbol{\beta}'\sigma_t^2. \quad (30)$$

A F-ARCH(1,1,1) model with  $n$  assets requires  $n(n+5)/2$  parameters. The model addresses both of the main issues running through the MGARCH literature since it only requires the estimation of univariate GARCH models for the  $k$  factors and  $\mathbf{H}_t$  is guaranteed to be positive definite. The Orthogonal GARCH (O-GARCH) model of Alexander and Chibumba (1997) and the Generalized Orthogonal (GO-GARCH) model of Van der Weide (2002) also rely on a linear transformation of unobserved factors that follow GARCH processes but instead assume that these factors are uncorrelated with one another. Returns are given by:

$$\mathbf{r}_t = \mathbf{W}\mathbf{f}_t, \quad (31)$$

where the invertible  $n \times n$  transformation matrix,  $\mathbf{W}_t$ , is estimated from the data and  $\mathbf{f}_t$  is the  $n \times 1$  vector of factors<sup>25</sup>. The  $n \times n$  matrix of the conditional variances for the factors are driven by a GARCH process:

$$\mathbf{H}_t^f = (\mathbf{I} - \mathbf{A} - \mathbf{B}) + \mathbf{A} \odot (\mathbf{f}_{t-1} \mathbf{f}_{t-1}') + \mathbf{B} \mathbf{H}_{t-1}^f, \quad (32)$$

where  $\odot$  defines the Hadamard product or element-by-element multiplication. The individual GARCH processes that the factor loadings follow can be from any of the aforementioned univariate families described earlier. The unconditional variance of each factor is normalized to 1, thus  $E(\mathbf{f}_t \mathbf{f}_t') = \mathbf{I}_n$  and the conditional covariance of returns is given by:

$$\mathbf{H}_t = \mathbf{W} \mathbf{H}_t^f \mathbf{W}'. \quad (33)$$

The difference between O-GARCH and GO-GARCH is that the matrix that transforms the data into a set of uncorrelated factors must be orthogonal in the former model but only invertible in the latter. While these models are particularly useful when  $n$  is large, we have nonetheless included the GO-GARCH model in the forecasting set in order to have a representative from each of the prominent MGARCH models, despite the fact that in our case  $n = 2$ . Furthermore, its not inconceivable that the exchange rate and the asset prices are driven by common factors such as the interest rate or GDP, and the usage of a factor model makes theoretical sense in our bivariate application.

### 2.4.3 Non-Linear Combinations of Univariate GARCH models

**Conditional Correlation Models** Bollerslev (1990) proposed a model with time-varying conditional covariances and variances but with Constant Conditional Correlations (CCC) thereby simplifying the estimation and inference procedures. In addition the conditions needed to ensure positive-definiteness of  $\mathbf{H}_t$  in the CCC model are easier to impose and verify. These models imply a two step procedure whereby one first selects a suitable GARCH process for each conditional variance and then obtains the conditional covariance using the correlations. The correlation between assets  $i$  and  $j$  is given by:

$$\rho_{i,j,t} = \frac{h_{i,j,t}}{\sqrt{h_{i,i,t}} \sqrt{h_{j,j,t}}}. \quad (34)$$

The conditional covariance matrix is decomposed into the correlation matrix  $\mathbf{R}$ , which is then pre- and post-multiplied by the diagonal matrix  $\mathbf{D}_t$ :

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t. \quad (35)$$

---

<sup>25</sup>In the GO-GARCH model it is not possible to have less factors than there are assets, unlike the O-GARCH model.



$$\mathbf{D}_t = \text{diag}\{\sqrt{h_{i,t}}\}, \quad (36)$$

where  $\mathbf{D}_t$  is a diagonal matrix consisting of the conditional standard deviations. Provided that  $\mathbf{R}$  is positive definite and the  $n$  conditional standard deviations are well defined, then  $\mathbf{H}_t$  will be positive definite for all  $t$ . The most appealing aspect of this formulation is that it only requires the estimation of  $n$  univariate GARCH models, before estimating the covariance matrix through (35) by setting  $\mathbf{R}$  equal to the sample correlation matrix. In Bollerslev (1990) the author utilizes a GARCH(1,1) specification to model the conditional variances of five major currencies, before and after the implementation of the European Monetary System in 1979, however, as the CCC models are flexible they could allow for different univariate volatility structures such as EGARCH(1,1) or GARCH(1,2). Indeed it is this flexibility, along with the reduced estimation complexity that are the key advantages of this methodology.

**Dynamic Conditional Correlation** The assumption of constant conditional correlations may be considered unrealistic in some or indeed most settings and Engle (2002a) as well as He and Teräsvirta (2004) generalized the CCC model by affording the conditional correlation matrix  $\mathbf{R}_t$  a dynamic structure. In practice the Dynamic Conditional Correlation (DCC) model is nothing more than a GARCH model applied to the standardized residuals which are then rescaled in order to obtain the correlation matrix. From Equation (3) we know that  $\mathbf{z}_t = \mathbf{D}_t^{-1} \mathbf{r}_t$ , this then implies that  $E_{t-1}(\mathbf{z}_t \mathbf{z}_t') = \mathbf{D}_t^{-1} \mathbf{H}_t \mathbf{D}_t^{-1} = \mathbf{R}_t$  as  $E_{t-1}(\mathbf{r}_t \mathbf{r}_t') = \mathbf{H}_t$ <sup>26</sup>. Another representation of the conditional correlation  $\mathbf{R}_t = [\rho_{i,j,t}]$  is:

$$\rho_{i,j,t} = \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}} \sqrt{q_{j,j,t}}}. \quad (37)$$

Or in matrix form:

$$\mathbf{R}_t = \bar{\mathbf{Q}}_t^{-1/2} \mathbf{Q}_t \bar{\mathbf{Q}}_t^{-1/2}, \quad (38)$$

where  $\mathbf{Q}_t$  is the covariance matrix of the standardized residuals with a typical element  $\mathbf{Q}_t = [q_{i,j,t}]$  and  $\bar{\mathbf{Q}}_t = \text{diag}(Q_t)$ .  $\mathbf{Q}_t$  can be parameterized in a number of ways. If  $q_{i,j,t}$  were to follow a GARCH(1,1) specification then the process for the conditional covariance would be described by:

$$q_{i,j,t} = \bar{\rho}_{i,j} + \alpha(z_{i,t-1} z_{j,t-1} - \bar{\rho}_{i,j}) + \beta(q_{i,j,t-1} - \bar{\rho}_{i,j}), \quad (39)$$

where  $\bar{\rho}_{i,j}$  is the unconditional correlation between  $z_{i,t}$  and  $z_{j,t}$ . Recall that  $\text{var}(z_i) = q_{i,i} = 1$ , which implies that the unconditional expectation of each term in the denominator of Equation (37) is 1 and the unconditional covariance is equivalent to the unconditional correlation:

$$\bar{q}_{i,j} = \bar{\rho}_{i,j}. \quad (40)$$

---

<sup>26</sup>Recall returns have zero mean.

In matrix notation the GARCH model is given by:

$$\mathbf{Q}_t = \mathbf{S}(1 - \alpha - \beta) + \alpha(\mathbf{z}_{t-1}\mathbf{z}_{t-1}') + \beta\mathbf{Q}_{t-1}. \quad (41)$$

$\mathbf{S}$  is the unconditional covariance matrix of the vector  $\mathbf{z}_t$  and is equivalent to the unconditional correlation matrix, as was shown in Equation (40). Therefore,  $\mathbf{Q}_t$  is a weighted average of a positive-definite matrix,  $\mathbf{S}(1 - \alpha - \beta)$  and a positive semi-definite matrix,  $\alpha(\mathbf{z}_{t-1}\mathbf{z}_{t-1}')$ .  $\mathbf{S}$  is made equal to the sample correlation matrix of  $\mathbf{z}_t$ . This is known as "Variance targeting"<sup>27</sup>. The GARCH(1,1) model requires  $(N + 1)(N + 4)/2$  parameters to be estimated and is relatively parsimonious when compared to the Full BEKK and VEC models. The DCC model also inherits the CCC model's flexibility with regards to the allowing different univariate GARCH models for each asset. Unlike VEC or BEKK, the DCC model parameterizes conditional correlations directly, and holds a significant computational advantage over these models as the number of parameters to be estimated in the correlation process is independent of the number of series that are correlated, given that the dynamics of each conditional correlation is driven by the scalars  $\alpha, \beta$ <sup>28,29</sup>.

## 2.5 Estimation

Parameter estimation for  $\theta$  is relatively straightforward once a distributional assumption for the standardized residuals,  $z_t$ , is made. Assuming  $z_t$  is normally distributed and by implication  $r_t|I_{t-1} \sim N(\mu_t, h_t)$ , the Maximum Likelihood Estimates (MLE)  $\hat{\theta} = (\hat{\mu}_t, \hat{\omega}, \hat{\alpha}, \hat{\beta})$  can be found by maximizing the product of the conditional densities  $f(r_t|I_{t-1}, \hat{\theta})$  using the history of returns in the information set,  $I_{t-1}$ . The log-likelihood function  $l(\theta|r_t)$  takes the expression:

$$l(\theta|r_t) = -\frac{1}{2} \left[ T \log(2\pi) + \sum_{t=1}^T \log(h_t) + z_t^2 \right] \quad (42)$$

---

<sup>27</sup>Conditional covariance models have estimated parameters increasing at an order greater than the number of assets. One means of tackling this issue is by "targeting" as proposed by Engle and Mezrich (1996) which entails setting the unconditional variance  $\sigma^2$  or the unconditional covariance matrix  $\Sigma$  equal to the sample variance or covariance. Formally, variance targeting takes place if the following two definitions are met, Caporin and McAleer (2009):

- The intercept is an explicit function of the long run covariance
- The long run covariance solution is replaced by a consistent estimator of the unconditional sample covariance of the observed data.

For the GARCH(1,1) model the long run variance is  $\bar{\sigma}^2 = \omega/(1 - \alpha - \beta)$ , and if  $\bar{\sigma}^2$  is replaced with the sample variance then the model is said to be "targeted". Only  $\alpha$  and  $\beta$  need be estimated using MLE and  $\omega$  can be calculated using the sample variance. The decline in computational complexity in the univariate case is dwarfed by that of the multivariate case and many MGARCH models have targeted versions that reduce estimation time and make higher dimensional applications feasible.

<sup>28</sup>Combining Equations (37) and (41) one can express the conditional correlation  $\rho_{i,j,t}$  as:

$$\rho_{i,j,t} = \frac{(1 - \alpha - \beta)q_{i,j} + \alpha z_{i,t-1}z_{j,t-1} + \beta q_{i,j,t-1}}{\sqrt{((1 - \alpha - \beta)q_{i,i} + \alpha z_{i,t-1}^2 + \beta q_{i,i,t-1})(1 - \alpha - \beta)q_{j,j} + \alpha z_{j,t-1}^2 + \beta q_{j,j,t-1}}}$$

<sup>29</sup>This also ensures that  $\mathbf{R}_t$  is positive definite for all  $t$ .

The log-likelihood expression is highly non-linear in the parameters and maximization is carried out using a numerical or iterative search algorithm, one example of which is the BHHH algorithm of Berndt et al. (1974). Provided that the model is correctly specified and a set of technical regularity conditions are met then these MLE estimates are efficient, unbiased, consistent and asymptotically normal which permits inference on the parameter vector by the way of some standard Wald or T-tests, (Engle et al. (1985)). While the standard assumption of normality for  $z_t$  in combination with time varying volatility leads to increased leptokurtosis (relative to the normal distribution) in the unconditional distribution of  $r_t$ , the empirical distribution of daily or weekly returns contains even greater leptokurtosis still, resulting in miss-specification of  $l(\theta|r_t)$  and a loss of efficiency. Fortunately these Quasi-Maximum Likelihood Estimates (QMLE) will still be unbiased, consistent and asymptotically normal provided that the conditional mean and conditional variance functions are correctly specified<sup>30</sup>. However inference will need to be carried out using robust standard errors or the "sandwich-form" of the covariance estimator<sup>31</sup>. More crucially, point estimates for the forecasted value of  $h_t$  do not rely on the distribution of  $z_t$ , and this means that asymptotically valid forecasts can be made from the QMLE  $\hat{\theta}$ .

One of the most common findings in the empirical finance literature is that of leptokurtosis in the distributions of returns, see Mandelbrot (1963) and Fama (1965). Leptokurtic distributions that have been considered include the power-exponential distribution in Baillie and Bollerslev (2002) and the normal-lognormal mixture distribution in Hsieh (1989). However, the two most popular alternatives to the normal density and the ones that will be considered in this paper, are the Student's-t distribution suggested by Bollerslev (1987) and the Generalized Error distribution (GED) of Nelson (1991). The Student's-t distribution has the following pdf:

$$f(e_t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\Gamma(\frac{\nu}{2})}(\nu-2)^{-\frac{1}{2}}(h_t)^{-\frac{1}{2}} \left[1 + \frac{e_t^2}{h_t(\nu-2)}\right]^{-\frac{(\nu+1)}{2}}, \quad (43)$$

where  $\Gamma(\cdot)$  is the Gamma Function and  $\nu$  the shape parameter which must be constrained to be greater than two in order for the second moment to exist. The GED has the following pdf:

$$f(e_t) = \frac{\nu \exp[-(\frac{1}{2}) \left| \frac{e_t}{\lambda h_t^{1/2}} \right|^\nu]}{h_t^{1/2} \lambda 2^{(1+\frac{1}{\nu})} \Gamma(\frac{1}{\nu})}, \quad (44)$$

with

$$\lambda = [(2^{-\frac{2}{\nu}} \Gamma(\frac{1}{\nu})) / \Gamma(\frac{3}{\nu})]^{\frac{1}{2}},$$

where  $\nu$  is again the shape parameter. Thus estimating the GARCH model using MLE and either of these distributions will augment the parameter vector  $\theta$  with  $\nu$ . The closer the selected distribution is to

---

<sup>30</sup>  $E_{t-1}(e_t) = 0$  and  $E_{t-1}(e_t^2) = h_t$

<sup>31</sup> See Bollerslev and Wooldridge (1992).

the actual distribution of the data, the more efficient are the estimates of  $\theta$ <sup>32</sup>. Issues of efficiency, aside if the goal is to use the model for estimating quantiles, such as the VaR, then these alternative distributions will provide better estimates of the tails of these forecasted densities and more accurate risk measures.

Parameter estimation when  $n > 1$  is complicated and slowed down significantly by the need to carry out repeated matrix inversions. Assuming that the vector of standardized residuals,  $\mathbf{z}_t$ , is multivariate normal the log-likelihood function  $l(\theta|\mathbf{r}_t)$  will be given by:

$$l(\theta|\mathbf{r}_t) = -\frac{1}{2} \sum_{t=1}^T \log |\mathbf{H}_t| - \frac{1}{2} \sum_{t=1}^T \mathbf{r}_t' \mathbf{H}_t^{-1} \mathbf{r}_t \quad (45)$$

Maximization is then carried out iteratively, and given the assumption of heteroskedasticity,  $\mathbf{H}_t$  has to be inverted at every  $t$ , a computationally demanding requirement particularly when  $T$  and  $n$  are large.

### 3 Realized Volatility

Merton (1980) first noted that by assuming a diffusion-type process for market returns, the volatility of an asset over a given period could be estimated more accurately using the sum of squared intra-period returns. This has been termed the "realized volatility" approach and it has two important implications for volatility modelling. Firstly, it has been suggested by Andersen and Bollerslev (1998) that the evaluation of volatility forecasts carries much more weight when an accurate measure of ex-post volatility, such as the realized volatility, is used. Secondly, that by calculating realized volatility and treating it as an observed process (as opposed to an unobserved latent process), the application of traditional time series techniques to the direct modelling of volatility becomes possible, (Andersen et al. (2001)). Each of these issues is discussed in turn below.

#### 3.1 Realized Volatility and forecast evaluation

When evaluating the out-of-sample forecasting performance of various GARCH models the norm within the early literature was to simply square the out-of-sample returns to obtain the forecast benchmark. This resulted in a consistent asymmetry between the in-sample and out-of-sample performance of these models. While this practice is certainly justified within the GARCH setting given that  $E_{t-1}[r_t^2] = E_{t-1}[\sigma_t^2 z_t^2] = \sigma_t^2$ , the standardized residual,  $z_t$ , has a large degree of variation relative to that of  $\sigma_t^2$ , which results in a non-negligible amount of measurement error, (Andersen and Bollerslev (1998)). Therefore, despite its

---

<sup>32</sup>Engle and Gonzalez-Rivera (1991) propose estimating the distribution non-parametrically. First the model is estimated under the Gaussian assumption and the density of the standardized residuals  $z_t$  is then estimated using a linear spline. This estimated density is then used to maximize a new likelihood function. Using Monte Carlo simulations the authors show efficiency gains relative to QMLE.

unbiasedness for the true variance, the squared return is a noisy estimator<sup>33</sup>. The effect of the noise in squared returns on forecast evaluation is highlighted in Andersen and Bollerslev (1998), who conduct Mincer-Zarnowitz<sup>34</sup> regressions and find that the same forecasting model has an  $R^2$  of 5% or approximately 50% depending on whether the daily return squared or the daily realized volatility is used as the object of interest, a vast discrepancy. Hansen and Lunde (2005) produce a very similar finding for a vast number of GARCH specifications.

The increased availability of high-frequency data has germinated numerous branches of volatility research. One such branch has focused on how best to utilize this high-frequency data to improve upon existing approaches to volatility measurement by obtaining better estimates of ex-post volatility for periods of lower frequency such as daily returns. The most common means of doing so amongst practitioners is to simply sum the squared returns of some higher frequency that fall within a particular period. This realized variance provides the natural benchmark for evaluating volatility forecasts since it is an unbiased and consistent estimate of the price variability over a given time interval<sup>35,36</sup>. Observed returns are viewed as discrete observations from an underlying continuous-time process. Let us denote the efficient market price that would prevail in the absence of any market microstructure noise as  $P(t)$ . Now suppose that the logarithmic price process,  $p(t) = \log P(t)$ , follows a continuous sample-path and has the following stochastic differential equation (SDE) form:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t), \quad (46)$$

where  $\mu(t)$  denotes the drift,  $\sigma(t)$  the instantaneous volatility/diffusion coefficient and  $W(t)$  a standard Brownian motion<sup>37</sup>. The one period return  $r(t+1)$  is then given by:

$$\begin{aligned} r(t+1) &= p(t+1) - p(t), \\ &= \int_t^{t+1} \mu(\tau)d\tau + \int_t^{t+1} \sigma(\tau)dW(\tau), \end{aligned} \quad (47)$$

where  $t < \tau < t+1$ . It can then be shown, using the theory of quadratic variation for semi-martingales, that the formal ex-post measure of return volatility is given by the integrated variance (IV), (Andersen et al. (2001), Barndorff-Nielsen and Shephard (2002))<sup>38</sup>:

<sup>33</sup>Both the univariate and multivariate models will require less noisy estimates of the variance or covariance if the out-of-sample evaluation is to be convincing.

<sup>34</sup>See Mincer and Zarnowitz (1969).

<sup>35</sup>This idea is not only restricted to high frequency data alone; French et al. (1987) for example use monthly volatility estimates constructed from daily returns to investigate the relationship between monthly stock returns and volatility.

<sup>36</sup>A comprehensive review of this topic is perhaps beyond the scope of this paper. For a rigorous treatment of this area, to grasp the diffusion-theoretic underpinnings, the interested reader can consult Andersen et al. (2002) for a recent survey.

<sup>37</sup>The SDE could include a jump component that allows for discontinuities in the price path that are often associated with macroeconomic news or data releases, (Andersen et al. (2007), Andersen et al. (2006)). The realized power variation and realized bipower variation of Barndorff-Nielsen and Shephard (2004) are robust to these jumps.

<sup>38</sup>In a more general setting, i.e. in the presence of a jump component this measure is termed the quadratic variation.

$$IV(t+1) = \int_t^{t+1} \sigma^2(\tau) d\tau. \quad (48)$$

If the drift and instantaneous volatility are stochastic, then we can only form expectations of the future volatility conditional on the information set,  $I_t$ . Over longer horizons, such as daily or weekly intervals, the variation in the mean or drift is negligible and therefore the conditional variance of returns is approximately:

$$\text{var}(r(t+1)|I_t) \approx E \left[ \int_t^{t+1} \sigma^2(\tau) d\tau \mid I_t \right] \approx E[IV(t+1) \mid I_t] \quad (49)$$

To formalize the realized variance estimator consider a trading period such as one day beginning at  $a$  and ending at  $b$ , and the interval  $[a, b]$  being divided into  $m$  subintervals such that the sampling frequency is defined as  $\Delta = \frac{b-a}{m}$ . The daily return,  $r_{t+1}$ , is then just the sum of  $m$  intraday returns, denoted  $r_{j,m}$ :

$$r_{t+1} = \sum_{j=1}^m r_{j,m}. \quad (50)$$

Let  $\sigma_{t+1}^2$  be the actual unobserved variance for the trading period at time  $t+1$ , then the "sum-of-squared returns" realized variance estimator may be given by:

$$\hat{\sigma}_{t+1}^2 = \sum_{j=1}^m r_{j,m}^2. \quad (51)$$

Thus  $\hat{\sigma}_{t+1}^2$  is just the sum of  $m$  intraday sample variances for the zero mean intraday returns. The realized variance estimator in (51) can be thought of as the discrete-time approximation to the IV in (49) which corresponds very closely to the conditional variance. To prove unbiasedness one needs to make the assumption that these intraday returns are conditionally Gaussian and i.i.d:

$$r_{j,m} \mid \sigma_t \sim n.i.d.(0, \sigma_{t+1}^2/m) \quad (52)$$

For  $1 \leq j \leq m$ . From Equations (50) and (52) it follows that  $r_{t+1} \sim N(0, \sigma_{t+1}^2)$ . Furthermore, Equation (52) implies that  $E[\hat{\sigma}_{t+1}^2] = \sigma_{t+1}^2$  given that  $E[r_{j,m}^2] = \text{var}(r_{j,m}) = \sigma_{t+1}^2/m$ . The important thing to note is that both unbiasedness and the convergence to the integrated variance hinge on the assumption that intraday returns are i.i.d.

Since the variance of the realized variance estimator is  $\text{var}(\hat{\sigma}_{t+1}^2) = \frac{2\sigma_{t+1}^4\Delta}{m}$ , the integrated variance with intraday prices that are sampled more and more frequently will lead to the realized variance converging on the IV, (Ait-Sahalia et al. (2005), Barndorff-Nielsen and Shephard (2002), Andersen and Bollerslev (1998))<sup>39</sup>. In reality though this limiting case is unattainable, since data constraints will limit the sample frequency. Moreover, the theoretical convergence and unbiasedness are not obtained in practice due to a smorgasbord of market microstructure features, such as: intraday periodic volatility patterns, non-synchronous trading effects, discretization of prices, and most glaringly the bid-ask spread which all violate the i.i.d assumption for

<sup>39</sup>More formally these authors have shown that  $p \lim_{\Delta \rightarrow 0} \hat{\sigma}_{t+1}^2 = \text{var}(r_{t+1}|I_t)$ .

returns underpinning these results. Concretely, market microstructure noise is taken to mean the deviation of observed or inferred prices from frictionless equilibrium prices, (Hansen and Lunde (2006), Aït-Sahalia et al. (2005)). This noise can introduce a significant amount of measurement error into the realized variance estimator, as has been suggested by Bandi and Russell (2008) and Hansen and Lunde (2006).

The most intuitive way to deal with these distortions would be to select a frequency low enough such that it is free of market microstructure effects, but still high enough to make use of the asymptotic results underlying its use, so as to provide an efficient estimate of the IV as suggested in Andersen et al. (2001) and Andersen et al. (2001). However, Aït-Sahalia et al. (2005) note that by arbitrarily selecting some sampling frequency, instead of making use of all available data, runs counter to the basic statistical principle that more data is generally better. Based on the findings of Bandi and Russell (2008) and Aït-Sahalia et al. (2005) the choice of the appropriate  $\Delta$  in the presence of microstructure noise is an empirical question and highly dependent on the market under consideration. Using the assumed process in (46), with zero mean/drift, allows us to describe the relation between the observed log price,  $\check{p}(j, m)$ , and the equilibrium log price,  $p(j, m)$ , as follows:

$$\check{p}(j, m) = p(j, m) + u(j, m),$$

where the noise component,  $u(j, m)$ , is covariance stationary with mean zero. The observed return for the  $j$ th intraday period is given;

$$\check{r}_{j,m} = r_{j,m} + u_{j,m} - u_{j-1,m}, \quad (53)$$

$$= r_{j,m} + e_{j,m}. \quad (54)$$

Thus in the presence of market microstructure noise, the observed returns follow an MA(1) process which introduces autocorrelation in the observed returns, since  $E[\check{r}_{j,m} \check{r}_{j-1,m}] \neq 0$ . The end result is that the realized variance estimator will yield the following biased measure in the presence of noise:

$$RV_t = \sum_{j=1}^m r_{j,m} + 2 \sum_{j=1}^m r_{j,m} e_{j,m} + \sum_{j=1}^m e_{j,m}^2. \quad (55)$$

Even if the returns are uncorrelated with the microstructure noise error, the proportion of the total return variance caused by the noise will increase as  $\Delta$  decreases (and  $m$  grows). In addition, the variance of the noise component will begin to dwarf that of the equilibrium price process, (Aït-Sahalia et al. (2005), Bandi and Russell (2008), Hansen and Lunde (2006)).

Aït-Sahalia et al. (2005) derive a closed-form expression that selects the optimal sampling frequency based on the process given in (46), but with no drift under the assumption that the noise is i.i.d. The optimized  $\Delta$  is a function of the variance of the noise component, the sample size, and the variance of actual returns, which effectively balances the trade-off between the bias and variance of the realized variance

estimator, vis-a-vis its MSE where the former is decreasing in  $\Delta$  and the latter increasing in  $\Delta$ . Bandi and Russell (2008) also address the bias/variance trade-off that is present during the selection of  $\Delta$ , and provide a parallel expression to that of Aït-Sahalia et al. (2005), that does not rely on a constant diffusion coefficient. Although not optimal in the statistical sense, a practical solution is to "pre-whiten" returns by filtering them. In this vein, Andersen et al. (2001) and Corsi et al. (2003) use an MA filter in order to create a series of residuals devoid of any lag-one autocorrelation.

A less formal approach is that of Andersen et al. (1999), who propose a graphical tool termed the "volatility signature plot" which aids in the selection of  $\Delta$ . The signature plot graphs the realized variance averaged over multiple days as a function of the sampling frequency  $\Delta$ . The bias induced by the bid-ask spread should dissipate as the sampling frequency decreases, which suggests that the signature plot may be used to select the highest frequency for which the average realized variance seems to have stabilized. The effect of market microstructure bias is readily apparent when observing volatility signature plots as the realized variance generally increases without bound as sampling frequency increases.

Other well established empirical regularities<sup>40</sup> in intraday financial data such as volatility clustering, excess kurtosis and deterministic intraday patterns, can also affect realized variance measures, but the main culprit in terms of bias is the autocorrelation in returns, (Hansen and Lunde (2006), Aït-Sahalia et al. (2005), Bandi and Russell (2008)). To this end, Bai et al. (2000) have measured any potential statistical gains to be made by using intraday data versus daily data, by calculating an Efficiency Ratio (ER) which is defined as:

$$ER(\hat{\sigma}_{intra,t}^2, \hat{\sigma}_{day,t}^2) = \frac{MSE(\hat{\sigma}_{intra,t}^2)}{MSE(\hat{\sigma}_{day,t}^2)},$$

where  $\hat{\sigma}_{intra,t}^2$  is the daily realized variance estimate obtained using intraday returns and  $\hat{\sigma}_{day,t}^2$  is the daily realized variance estimate using daily returns, therefore a better estimator would be one where  $ER(\hat{\sigma}_{intra,t}^2, \hat{\sigma}_{day,t}^2) < 1$ . The authors then compare two non-parametric realized variance estimators; that described in Equation (51) and another estimator from French et al. (1987), that allows for some serial dependence in the intraday returns. Using a year's worth of half-hour returns ( $\Delta = 30$ ) to estimate the daily realized variance using Equation (51), they find that the  $ER > 1$  when there is some degree of lag-one autocorrelation. The authors also note the intraday estimator's lack of precision in the presence of kurtosis, time-varying volatility and the deterministic patterns in the variance.

The empirical evidence for autocorrelation in exchange rate returns is particularly strong as numerous studies of the intraday FX market have shown there to be negative first-order correlation. Goodhart and Figlioli (1991) estimate the first order autocorrelation to be around -0.18 for one minute DMUSD returns measured over a three day period in 1987. Similarly, Andersen and Bollerslev (1997) use the same currency pair but using 5, 10, 20, 30 minute returns over the period October 1992 to September 1993, find autocorrelations of -0.04, -0.07, -0.082, and -0.043 respectively. More recently, Corsi et al. (2003) find lag-one autocorrelation as high as -0.4 for the USDJPY and USDCHF for tick data. For our purposes the selection

<sup>40</sup>See Baillie and Bollerslev (1991), Andersen and Bollerslev (1997), Muller et al. (1990), Goodhart and Figlioli (1991).



of an estimator that can effectively deal with autocorrelation will be critical, as we are dealing with relatively high-frequency data.

### 3.2 Estimators that are robust to market microstructure noise.

The detrimental impact of market microstructure noise on the statistical properties of the realized variance has spurred research into estimators that are robust to this noise. One such class of estimators are the kernel-based realized variance estimators or "realized kernels" which bear more than a passing resemblance to the HAC estimator of Newey and West (1987) which is used to estimate the long run variances and covariances of stationary stochastic processes. These estimators are initially presented in the context of the following two assumptions. Firstly, the equilibrium log price is a stochastic volatility martingale with no drift;  $dp(t) = \sigma(t)dW(t)$ , and secondly, the microstructure noise is independent of the equilibrium price, as well as being i.i.d with mean zero. Following Bandi and Russell (2011), the family of asymmetric realized kernel estimators can then take the following form:

$$RV_{AK} = w_0\hat{\gamma}_0 + 2\sum_{h=1}^q w_h(\hat{\gamma}_h), \quad (56)$$

where  $\hat{\gamma}_h = \sum_{j=1}^{m-h} r_j r_{j+h}$  is the  $h$ -step autocovariance function for  $h = 0, \dots, m-1$  and  $w_h$  is the kernel weighting function applied to the covariances that number  $q$  in total. Therefore, the realized kernel estimates the sample autocovariance and corrects the realized variance estimator for the autocorrelation that accompanies market microstructure noise. The first estimator of this kind was that of Zhou (1996), which corresponds to the case where  $w_0 = 1$ ,  $w_1 = \frac{m}{m-1}$  and  $w_h = 0$  for  $h \geq 2$ . This estimator only corrects for the first-order autocorrelation, which is more plausible for heavily traded markets and where  $\Delta$  is not less than 1 minute, (Barndorff-Nielsen et al. (2004)). However, Bandi and Russell (2011) note that the selection of  $q$ , or rather the ratio of  $q$  to the number of intra-period observations  $m$ , denoted  $\phi = \frac{q}{m}$ , has a significant impact on the small and large sample properties of these estimators. Hansen and Lunde (2004) follow a more general approach than that of Zhou (1996) and correct for the first  $q_m$  autocorrelations where  $q_m$  is increasing in  $m$  ( $\phi$  is fixed). They make use of Bartlett-type kernel weights where  $w_0 = \frac{m-1}{m} \frac{q-1}{q}$  and  $w_h = \frac{q-h}{q}$  for  $h = 1, \dots, q$ . The class of asymmetric kernel estimators are unbiased but inconsistent, (Barndorff-Nielsen et al. (2004), Bandi and Russell (2011)).

A close relative to the asymmetric kernel is the "Two Time Scales Estimator" (TTSE) of Zhang et al. (2005). Barndorff-Nielsen et al. (2004) note that this sub-sampling method is essentially a modified Bartlett-type kernel estimator. The TTSE entails partitioning the grid of observation times within the trading period into non-overlapping subgrids, calculating realized variance for each of those subgrids and then averaging them. This is then combined with a scaled version of the regular realized variance to produce an unbiased and consistent estimator of the integrated variance under the assumption that the noise is i.i.d. This estimator has been improved upon in Ait-Sahalia et al. (2011), to deal with the case where the microstructure noise

is not i.i.d.

More recently Barndorff-Nielsen et al. (2008) have proposed flat-top symmetric realized kernels of the following form:

$$RV_K = \hat{\gamma}_0 + \sum_{h=1}^q w_h (\hat{\gamma}_h + \hat{\gamma}_{-h}), \quad (57)$$

where  $\hat{\gamma}_h = \sum_{j=1}^m r_j r_{j-h}$  and  $w_h = k(\frac{h-1}{q})$ . The function  $k(\cdot)$  is defined on the interval  $[0, 1]$ , such that  $k(0) = 1$  and  $k(1) = 0$ . Kernels satisfying these criteria include the Parzen, Tukey-Hanning and Cubic kernel. The estimator of Barndorff-Nielsen et al. (2008) is both unbiased and consistent, regardless of whether the noise is correlated with itself or the efficient prices<sup>41</sup>. Using this framework, in much the same way as the optimal realized variance estimator can be selected based on the appropriate sampling frequency  $\Delta$ , the optimal realized kernel will minimize the MSE based on the right selection of  $\phi$ , (Bandi and Russell (2011)).

Another distinct approach that can handle microstructure noise, but one that is closely related to the "pre-whitening" techniques of Andersen et al. (2001), is that put forward by Hansen et al. (2008). Hansen et al. (2008) has shown that realized variance, calculated using MA(1) filtered returns and scaled appropriately, yield an unbiased and consistent estimate of the integrated variance, provided that the noise is i.i.d and independent of the efficient prices. The MA(1) based realized variance estimator is given by:

$$RV_{MA} = (1 - \hat{\theta})^2 \sum_{i=1}^m \hat{e}_{i,m}^2, \quad (58)$$

where  $\hat{\theta}$  is the MA(1) estimate and  $\hat{e}_{i,m}^2$  are the filtered residuals. This estimator is based on the insight that the MA(1) filtered residuals still possess unwanted variance from the microstructure noise and this needs to be accounted for by scaling the residuals. The approach prior to Hansen et al. (2008) had been to calculate realized variance as the sum of the squared MA filtered residuals, and while this method does provide information regarding the dynamic behavior of the IV, it is unsuitable for estimating the level of IV<sup>42</sup>.

Note that the  $RV_{MA}$  estimator incorrectly assumes that the intraday volatility is constant, however, the miss-specified  $RV_{MA}$ 's performance in a simulation study with non-constant volatility<sup>43</sup> and four different types of microstructure noise, was mostly superior or at least on par with the kernel estimators of Zhou (1996) and Hansen and Lunde (2006), as well as the sub-sampling estimators of Zhang et al. (2005) and Aït-Sahalia et al. (2011)<sup>44</sup>.

---

<sup>41</sup>Note that the TTSE estimator of Zhang et al. (2005) requires that the noise be independent on both counts in order for unbiasedness and consistency to be established. For a summary of the statistical properties of the asymmetric and symmetric kernels as well the TTSE estimator, see McAleer and Medeiros (2008).

<sup>42</sup>Evidence of this is provided in Andersen et al. (2001) where the unscaled MA(1) based estimator overstates the IV by 62%.

<sup>43</sup>We refrain from labelling the volatility as stochastic since it was deterministically specified in such a way as to simulate the familiar U-shape volatility observed over the course of the trading day in equity markets.

<sup>44</sup>Two MA-based estimators were considered in the simulation, these included an MA(1) and an MA(20), with the former being pitted against other estimators designed to handle i.i.d noise, those of Zhou (1996) and Zhang et al. (2005), while the

Under each noise specification; i.i.d noise, rounding noise as well positive and negative correlation with the efficient price, the MA-based estimators generally outperformed or matched the aforementioned estimators on the basis of MSE. If there was higher order autocorrelation present, which is usually the case at higher sampling frequencies, then the MA(20) is superior to the MA(1) estimator. Interestingly, for sampling frequencies greater than 120 seconds, the MSE's of each estimator were virtually identical, suggesting that the selection of an estimator is somewhat arbitrary at low frequencies such as five minutes. However if the higher order dependencies are close to zero, then the MA(20) estimator will estimate unnecessary coefficients, which in turn reduces the accuracy of the estimator.

Based on the above discussion, the choice between the different "noise-robust" estimators rests on the assumptions made with regards to the noise; specifically whether the noise is correlated with itself and/or the efficient prices. In this regard, Hansen and Lunde (2006) show intertemporal dependence, but only when the data are sampled at ultra high frequencies such as one tick. Market structure largely determines whether the noise is independent of the efficient price, (Bandi and Russell (2006)). For a decentralized market where agents are spread out and place trades based on differing opinions of value, the random price changes that occur as a result of these trades should be uncorrelated with the underlying efficient price, Bandi and Russell (2006). The FX market clearly meets this description as trading takes place through the decentralized interbank market where multiple agents post quotes based on their inventory and perception of value.

### 3.3 Realized Volatility for Forecasting

The traditional GARCH methodology does not incorporate high frequency returns when forming predictions of daily volatility. Over the last 25 years, the availability of data for currencies and equities sampled at frequencies higher than daily has increased substantially<sup>45</sup> and a significant area of research has been aimed at the modelling of volatility directly. Noteworthy examples of this methodology are provided in Andersen et al. (2001), Andersen et al. (2001), Andersen et al. (2003) and Andersen et al. (2007); all of which make use of reduced-form time series models to forecast the realized variance or some transformation of the realized variance. For example, Andersen et al. (2001) model the DMUSD and JPYUSD exchange rates using a standard linear Gaussian VAR and find evidence for long memory in the realized variance process. Andersen et al. (2003) go one step further and model these exchange rates along with the JPYDM rate using these insights as a guide. They find that the unconditional distributions of the realized standard deviation for the DMUSD, JPYUSD as well as the implied JPYDM are skewed to the right and leptokurtic but that the skewness and kurtosis for the logarithmic realized standard deviations are very close to that latter was compared against other estimators that could handle deviations from the i.i.d assumption those of Hansen and Lunde (2006) and Art-Sahalia et al. (2011).

<sup>45</sup>In terms of FX prices the database of Olsen and Associates of Zurich (O&A) has been widely used while the Trade and Quotations (TAQ) database contains intraday prices for equities traded on the main US exchanges.

of the normal distribution<sup>46</sup>. They also find evidence that the log-realized variance possesses dynamics indicative of a long memory process. This result taken together with the log-normality of the realized standard deviation motivates their implementation of a trivariate long memory Gaussian VAR (5)<sup>47</sup>. The out-of-sample forecasting ability of the model is then compared to a number of competing models such as the FIEGARCH model of Bollerslev and Ole Mikkelsen (1996), a standard GARCH(1,1) model as well as J.P. Morgan's RiskMetrics model. The long memory VAR's one-day-ahead forecasts are found to be far superior to the competing models, when using the Mincer-Zarnowitz type regressions to evaluate them<sup>48</sup>.

While research into the direct modelling of volatility has proven fruitful another branch of literature has sought to fuse the traditional GARCH methodology with the new generation of improved volatility measures, thereby providing greater informational content and more accurate forecasts. Models that have some measure of realized volatility as an exogenous right-hand-side variable for the conditional variance have been labelled GARCH-X models. An example is given in Engle (2002b) who make use of a Multiplicative Error Model (MEM)<sup>49</sup> to incorporate the realized variance into a GARCH(1,1) model. The problem with GARCH-X models is that they are agnostic with regards to the dynamics of the realized variance measures and cannot produce multi-step forecasts. The approach of Engle and Gallo (2006) builds on the MEM of Engle (2002b) and utilizes three different indicators of volatility; the squared daily returns, the squared daily high/low range as well as the realized variance. All three measures follow an MEM structure and the introduction of a GARCH process for the conditional mean of each measure essentially closes the model. Using S&P500 data the model produces forecasts that capture persistence in equity market volatility as measured by the VIX index, although no comparisons are made between the MEM model forecasts and those of standard GARCH models. The Realized GARCH or RealGARCH model of Hansen et al. (2012) specifies a more general ARMA structure for the realized volatility measure<sup>50</sup>, which they term the measurement equation. The model has a linear or log-linear specification and both models were able to provide a superior empirical fit for 28 US stocks and the exchange-traded S&P 500 fund.

Another recent introduction to this branch of literature is the HEAVY or High-frEQUENCY-bAsed Volatility model of Shephard and Sheppard (2010) which continues in the same vein but reverts to the GARCH methodology of Bollerslev (1986). In simple terms, the univariate HEAVY model makes use of a measure of

<sup>46</sup>Interestingly, these results are not confined to the currencies space. Andersen et al. (2001) find that the unconditional distributions for the realized volatility and the logarithmic realized volatility for the 30 stocks that make up the Dow Jones Industrial Index (DJIA) are also non-normal/skewed and normal respectively. They also find evidence to suggest a fractionally integrated process.

<sup>47</sup>The lag length was arbitrarily chosen as the number of days in a week.

<sup>48</sup>In a similar vein, Andersen et al. (2007) make use of the realized bi-power variation of Barndorff-Nielsen and Shephard (2004) and an AR model to generate superior volatility forecasts.

<sup>49</sup>The MEM model is used specifically in the context of non-negative time series. The standard MEM specification is given by:  $x_t = u_t e_t$ ,  $e_t | I_{t-1} \sim D(1, \phi_t^2)$

Where  $e_t$  is a non-negative disturbance term and  $u_t$  the mean. The original ARCH model of Engle (1982) is in fact an MEM model since  $r_t^2 = h_t e_t^2$

<sup>50</sup>In that particular paper the authors make use of the realized kernel estimator of Barndorff-Nielsen et al. (2008).

the realized variance from the most recent period to obtain its conditional variance forecast instead of the prior period's squared innovations. Recall our description of the GARCH(1,1) process in Equation (8):

$$h_t = \omega + \alpha(e_{t-1})^2 + \beta(h_{t-1}). \quad (59)$$

The specification of a HEAVY(1,1) process for the conditional variance would take the form:

$$\text{var}(r_t | I_{t-1}^{HF}) = h_t = \omega_H + \alpha_H(RM_{t-1}) + \beta_H(h_{t-1}), \quad (60)$$

where  $I_{t-1}^{HF}$  consists of the high-frequency returns  $r_{j,m}$ , as well as the past realized variance measures. The measure of the realized variance  $RM$ , can be any of the estimators discussed in the prior section. For example, Shephard and Sheppard (2010) make use of the realized-kernel estimator of Barndorff-Nielsen et al. (2008). If  $\mu = 0$  then  $(e_{t-1})^2 = (r_{t-1})^2$ , so based on the discussion above regarding volatility measurement, the main distinction between a standard GARCH model and the HEAVY model becomes clear; the HEAVY model relies on a less noisy estimator of the previous day's volatility than the GARCH model. The system also includes a process for  $RM_t$ , although this will only be needed if the purpose of the exercise is to form multi-step forecasts for  $h_t$ , which obviously require multi-step forecasts for the realized variance:

$$E(RM_t | I_{t-1}^{HF}) = v_t = \omega_R + \alpha_R RM_{t-1} + \beta_R v_{t-1}, \quad (61)$$

where  $\alpha_R$  is the AR parameter for  $RM_t$ . Equation (60) then models the close-to-close conditional variance and is referred to as the HEAVY-r model, while Equation (61) models the conditional expectation of the open-to-close variance and is referred to as the HEAVY-RV model<sup>51</sup>.

Equation (11) showed that in the GARCH(1,1) model, the conditional variance reverted monotonically back towards its long-run mean as the forecast horizon increased, HEAVY models allow for additional momentum; in that the conditional variance may continue on an upward or downward trend before eventually reverting towards its long run equilibrium value<sup>52</sup>.

<sup>51</sup>A typical estimate for  $\beta_H$  is around 0.6-0.7. In contrast,  $\hat{\beta}$  is typically greater than 0.91, which indicates that the GARCH model has more memory in that it includes distant innovations.

<sup>52</sup>The HEAVY model has been extended to deal with multiple assets, (Noureldin et al. (2011)). Defining the outer product of returns as  $\mathbf{P}_t = \mathbf{r}_t \mathbf{r}_t'$  and the realized variance matrix as  $\mathbf{RM}_t$  then the covariance matrix and the expected realized covariance matrix  $\mathbf{M}_t$  can be given by:

$$E[\mathbf{P}_t | I_{t-1}^{HF}] = E[\mathbf{r}_t \mathbf{r}_t' | I_{t-1}^{HF}] = \mathbf{H}_t$$

$$E[\mathbf{V}_t | I_{t-1}^{HF}] = \mathbf{M}_t$$

Both the conditional covariance matrix and  $\mathbf{M}_t$  have been parameterized in a way similar to that of the BEKK model of Engle and Kroner (1995). If we assume a HEAVY(1,1) model then;

## 4 Forecasting Methodology

For practitioners the one-day-ahead forecast of the second moment of asset returns carries significantly more importance than the  $h$ -step forecast where  $h$  is typically 5, 10, or 22 days. Therefore we restrict our focus to a series of one-step-ahead forecasts generated from our respective models, which are then evaluated using various rigorous statistical measures. While out-of-sample forecast evaluation has been traditionally conducted by statistical measures such as the Mean Square Error (MSE), the most obvious means of comparison amongst models and one that could result in different rankings, would be to calculate the utility gained by the end-user of the forecasts, (West et al. (1993)). Univariate forecasts for exchange rates are mostly at odds with meaningful economic forecast evaluation. For example, a volatility timing strategy, such as the one examined in Fleming et al. (2003), requires expected returns that are not zero, since a rational investor would not allocate any of his wealth to a volatile asset (such as hard currency holdings), without any expected return. In this paper we will evaluate both the univariate and multivariate forecasts by statistical measures outlined below, but within the multivariate setting currency hedging will provide an additional means of comparing forecasts.

The main challenge encountered when evaluating volatility forecasts is that the forecasted variable is not observed. Researchers have addressed this problem using one of two approaches. The first employs statistical methods that make use of a proxy for the latent variable, where (51) and (80) are two examples of such proxies. The second method makes uses of economic evaluations to compare forecasts on the basis of mean-variance portfolio decisions, the portfolio's tracking error or the gains from hedging, (Patton and Sheppard (2009)).

### 4.1 Statistical evaluation of volatility forecasts

To set the stage for this discussion, we first introduce the statistical loss function  $L(\hat{\sigma}_t^2, h_t | I_{t-1})$ , which maps the ex-ante forecast (the conditional variance  $h_t$ ) and the ex-post realization (the realized variance estimate  $\hat{\sigma}_t^2$ ) into a loss value. Intuitively it must be the case that the optimal forecast  $h_t^*$  satisfies the following first order condition (FOC):

$$E_t \left[ \frac{\partial L(\hat{\sigma}_t^2, h_t^* | I_{t-1})}{\partial h_t^*} \right] = 0.$$

---

The optimal forecast vis-a-vis the loss function is the true variance;  $h_t^* = \sigma_t^2$ .

$$\begin{aligned} \mathbf{H}_t &= \mathbf{C}_H \mathbf{C}_H' + \mathbf{B}_H \mathbf{H}_{t-1} \mathbf{B}_H' + \mathbf{A}_H \mathbf{V}_{t-1} \mathbf{A}_H' \\ \mathbf{M}_t &= \mathbf{C}_M \mathbf{C}_M' + \mathbf{B}_M \mathbf{M}_{t-1} \mathbf{B}_M' + \mathbf{A}_M \mathbf{V}_{t-1} \mathbf{A}_M' \end{aligned}$$

As was the case with the BEKK model  $\mathbf{A}$  and  $\mathbf{B}$  could be restricted to be diagonal or scalar in order to increase the degrees of freedom and make estimation more feasible.

Realizations of the partial derivative above should on average be zero, or to put it another way, the forecasts should have uncorrelated prediction errors, (Andersen et al. (2006)). The obvious way of testing this would be to use a linear regression of the general form:

$$\frac{\partial L(\hat{\sigma}_t^2, h_t^* | I_{t-1})}{\partial h_t^*} = a + \mathbf{b}' \mathbf{x}_t + e_t,$$

where  $\mathbf{x}_t$  is a vector of possible explanatory variables observable at time  $t$  and  $\mathbf{b}$  the vector of coefficients. In order for the prediction errors to be zero, on average, it must be the case that  $a = b = 0$ , and this can be tested using  $t$  and/or  $F$ -tests. In the particular case of the quadratic loss function,  $L(\hat{\sigma}_t^2, h_t | I_{t-1}) = (\hat{\sigma}_t^2 - h_t)^2$ , the FOC is  $\frac{\partial L(\hat{\sigma}_t^2, h_t | I_{t-1})}{\partial h_t} = -2(\hat{\sigma}_t^2 - h_t)$ , and the regression should take the form:

$$(\hat{\sigma}_t^2 - h_t) = a + \mathbf{b}' \mathbf{x}_t + e_t.$$

Perhaps the most common regression of the form above is that of the Mincer and Zarnowitz (1969) hereafter (MZ), variety where the forecast is the only explanatory variable:

$$\hat{\sigma}_t^2 = \alpha + \beta h_t + e_t, \tag{62}$$

where  $H_0 : \alpha = 0, \beta = 1$  and  $H_1 : \alpha \neq 0, \beta \neq 1$ . Note that in this case the true regressand in the MZ regression is unobservable and we are forced to make use of a proxy. This makes the selection of a realized volatility estimator of key importance. The less accurate the volatility proxy the less accurate the estimates of  $\alpha$  and  $\beta$  which reduces the power of the test to reject the null, (Patton and Sheppard (2009))<sup>53,54,55</sup>.

#### 4.1.1 Robust Loss Functions

The  $R^2$  obtained from the regression in (62) is equivalent to the Mean Squared Error(MSE) metric which falls into a broader set of statistical loss functions that are used for forecast evaluation. The MSE can be calculated in the following way:

$$MSE = \frac{\sum_{t=1}^T (\hat{\sigma}_t - h_t^{1/2})^2}{T}. \tag{63}$$

---

<sup>53</sup>The errors from (62) will have heteroskedastic errors under the null hypothesis so estimation of the covariance matrix of the OLS parameters must be carried out using a heteroskedasticity consistent estimator, Andersen et al. (2006).

<sup>54</sup>Patton and Sheppard (2009) propose another means of dealing with the heteroskedastic errors, namely by using the volatility forecasts themselves to estimate the model using Generalized Least Squares (GLS).

<sup>55</sup>Multivariate forecast tests are generalized in a straightforward manner. One solution is to simply estimate a MZ regression for each unique element of  $\mathbf{H}_t$ . Alternatively, a test of the joint hypothesis that all the coefficients are zero can be run by vectorizing the process.

$$vech(\mathbf{\Sigma}_t) = \boldsymbol{\alpha} + diag(\boldsymbol{\beta})vech(\mathbf{H}_t) + \mathbf{e}_t,$$

where the  $vech()$  operator performs the now familiar stacking function of the lower triangular part of the matrix, and the  $diag()$  converts a  $n \times 1$  vector into a  $n \times n$  matrix, with the elements of  $\boldsymbol{\beta}$  on the diagonal.

Another popular measure in the literature and one that will be used in this paper is the QLIKE put forward by Bollerslev et al. (1994) which corresponds to the loss function implied by a Gaussian likelihood:

$$QLIKE = \frac{\sum_{t=1}^T (\log h_t + \hat{\sigma}_t^2/h_t)}{T}. \quad (64)$$

Incidentally, both of these loss function share a property that makes them desirable for out-of-sample forecast evaluation. Using only a volatility proxy, and not the actual latent volatility, necessitates that any loss function be "robust" to the usage of such a proxy, as the ranking of forecasts based on this proxy may not be the same asymptotic ranking achieved under the true volatility, (Andersen et al. (2005)). The optimal forecast,  $h_t^*$ , for any loss function, robust or not, must meet the following condition:

$$h_t^* = \arg \min_{h \in H} E_{t-1}[L(\hat{\sigma}_t^2, h_t)].$$

Patton (2011) formally defines a loss function as robust if the ranking of any two volatility forecasts  $h_t^a$  and  $h_t^b$  by their losses is invariant to the usage of the true conditional variance or a conditionally unbiased volatility proxy. In other words, for a loss function to be robust it must be the case that the optimal forecast vis-à-vis the loss function, is always the true conditional variance, regardless of whether the proxy  $\hat{\sigma}_t^2$  is used, or the true latent variance  $\sigma_t^2$ . More formally:

$$\begin{aligned} E_{t-1}[L(\hat{\sigma}_t^2, h_t^a)] &\geq E_{t-1}[L(\hat{\sigma}_t^2, h_t^b)] \\ \iff E_{t-1}[L(\sigma_t^2, h_t^a)] &\geq E_{t-1}[L(\sigma_t^2, h_t^b)] \end{aligned}$$

Patton (2011) provides necessary and sufficient conditions for a robust loss function and then derives a family of robust and homogenous<sup>56</sup> loss functions within which the MSE and QLIKE are nested:

$$\begin{aligned} L(\sigma_t^2, h_t; b) &= \frac{1}{(b+1)(b+2)}(\hat{\sigma}_t^2 - h_t^{b+2}) - \frac{1}{b+1}h_t^{b+1}(\hat{\sigma}_t^2 - h_t) \quad , \quad b \neq -1, -2, \\ &= h_t - \hat{\sigma}_t^2 + \hat{\sigma}_t^2 \log \quad , \quad b = -1, \\ &= \frac{\hat{\sigma}_t^2}{h_t} - \log \frac{\hat{\sigma}_t^2}{h_t} - 1 \quad , \quad b = -2. \end{aligned} \quad (65)$$

The scalar  $b$  captures the possible asymmetric effect of under- versus over-prediction. Symmetric loss functions such as the MSE have  $b = 0$  and punish over and under-prediction of volatility equally, in contrast the QLIKE (where  $b = -2$ ) punishes under-prediction more harshly<sup>57</sup>.

One way of measuring the distortion caused by using a noisy volatility proxy is to calculate the degree of bias in the optimal forecast under a given loss function when using such a proxy. A multiplicative bias term of one indicates that the loss function is unbiased. Patton and Sheppard (2009) conduct a simulation study with three possible DGPs; GARCH diffusion, log-normal diffusion and two-factor diffusion<sup>58</sup>. Nine possible

<sup>56</sup>Functions that are invariant to any re-scaling of returns.

<sup>57</sup>Patton and Sheppard (2009) show that the power of DM tests using the QLIKE loss function is higher than those using the MSE, suggesting preference over the QLIKE in pair-wise comparisons at a minimum.

<sup>58</sup>See Andersen et al. (2002).



loss functions were evaluated, including the MSE and QLIKE measures. For RV calculated using 30 and 5 minute returns, the amount of multiplicative bias for MSE and QLIKE, unlike any of the loss functions, was 1 for each of the simulated DGPs. This indicates that only these loss functions are robust in the sense described above, provided however that the volatility proxy is unbiased.

As pointed out by Bollerslev et al. (1994) there is no unique criterion by which to judge out-of-sample forecasts. Hansen and Lunde (2005) address this concern by evaluating forecasts in terms of seven different loss functions. Mindful of the findings of Patton (2011), the approach we adopt here is to only evaluate forecasts using robust loss functions.

#### 4.1.2 Multivariate loss functions

The need for a set of loss functions that are robust to the usage of a proxy is also present when one considers forecasts of the latent covariance matrix. Laurent et al. (2012) provide parallel multivariate conditions to those of Patton (2011) that describe a class of loss functions that produce the same ranking of covariance forecasts, whether the true conditional covariance matrix is used or an unbiased proxy of that matrix. This section will rely on an ex-post measure of the covariance. Fortunately, the realized covariance is a straightforward generalization of the realized variance estimator in (51) and can be obtained by summing the cross products of  $m$  intraday returns, sampled at an appropriate frequency, (Andersen et al. (2003)):

$$\hat{\Sigma}_t = \sum_{j=1}^m \mathbf{r}'_{j,m} \mathbf{r}_{j,m}. \quad (66)$$

Where  $\hat{\Sigma}_t$  is an  $n \times n$  matrix with the realized variance for assets  $i, \dots, n$  appearing on the diagonal while the realized covariance between asset  $i$  and asset  $j$  is given by  $\hat{\Sigma}_{t,i,j}$ <sup>59</sup>.

Analogous multivariate expressions that define the robustness property of multivariate loss functions are given by:

$$E_{t-1}[L(\hat{\Sigma}_t, \mathbf{H}_t^a)] \geq E_{t-1}[L(\hat{\Sigma}_t, \mathbf{H}_t^b)]$$

$$\iff E_{t-1}[L(\Sigma_t, \mathbf{H}_t^a)] \geq E_{t-1}[L(\Sigma_t, \mathbf{H}_t^b)].$$

Multivariate loss functions essentially measure the distance between the conditional covariance and realized covariance matrices in  $n$ -dimensional space. Under a certain set of assumptions, Laurent et al. (2012) present a family of robust and homogenous multivariate loss functions that are amongst the class of Bregman distances<sup>60</sup>. These loss functions are a function of the forecast errors,  $\hat{\Sigma}_t - \mathbf{H}_t$ , and take the following quadratic form:

<sup>59</sup>Andersen et al. (2003) show that if returns are linearly independent and  $n < m$ , then  $\hat{\Sigma}_t$  will be positive definite.

<sup>60</sup>See Bregman (1967).

$$L(\hat{\Sigma}_t, \mathbf{H}_t) = \text{vech}(\hat{\Sigma}_t - \mathbf{H}_t)' \hat{\Lambda} \text{vech}(\hat{\Sigma}_t - \mathbf{H}_t), \quad (67)$$

where  $\hat{\Lambda}$  is an  $n(n+1)/2 \times n(n+1)/2$  matrix of weights which determines whether covariance forecast errors and variance forecast errors are treated symmetrically or not. In the event that  $\hat{\Lambda} = \mathbf{I}_{n(n+1)/2}$  then the loss function is simply the Euclidean distance, where variances and covariances are equally weighted. A loss function that assigns double weights to covariance forecast errors is the Frobenius distance, where  $\hat{\Lambda} = \text{diag}(\text{vech}(\mathbf{V}))$  with  $v_{i,j} = 1$  if  $i = j$  and  $v_{i,j} = 2$  if  $i \neq j$  for  $i, j = 1, \dots, n(n+1)/2$ . For example, the Frobenius loss function for  $n = 2$  will take the form<sup>61</sup>:

$$L(\hat{\Sigma}_t, \mathbf{H}_t) = \begin{bmatrix} \hat{\Sigma}_{t,i,i} - \mathbf{H}_{t,i,i} & \hat{\Sigma}_{t,j,j} - \mathbf{H}_{t,j,j} & \hat{\Sigma}_{t,i,j} - \mathbf{H}_{t,i,j} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \hat{\Sigma}_{t,i,i} - \mathbf{H}_{t,i,i} \\ \hat{\Sigma}_{t,j,j} - \mathbf{H}_{t,j,j} \\ \hat{\Sigma}_{t,i,j} - \mathbf{H}_{t,i,j} \end{bmatrix}. \quad (68)$$

#### 4.1.3 Tests of Predictive Ability

The significance of any difference in the MSE or QLIKE between competing models is typically tested on a pair-wise basis using the test of Diebold and Mariano (1994) (hereafter, DM), where the null is that of Equal Predictive Ability (EPA) amongst the two models. If we define the series of forecasts from each model as  $\{h_t^a\}_{t=1}^T$  and  $\{h_t^b\}_{t=1}^T$ , then there are two possible alternative hypotheses to the null of EPA:

$$\begin{aligned} H_0 &: E[L(\hat{\sigma}_t^2, h_t^a)] = E[L(\hat{\sigma}_t^2, h_t^b)] \\ H_1 &: E[L(\hat{\sigma}_t^2, h_t^a)] > E[L(\hat{\sigma}_t^2, h_t^b)] \\ H_2 &: E[L(\hat{\sigma}_t^2, h_t^a)] < E[L(\hat{\sigma}_t^2, h_t^b)], \end{aligned} \quad (69)$$

The performance of model (a) relative to model (b) is given by:

$$d_t = L(\hat{\sigma}_t^2, h_t^a) - L(\hat{\sigma}_t^2, h_t^b). \quad (70)$$

The test statistic is computed as a regular  $t$ -test:

$$DMW_T = \frac{\sqrt{T} \bar{d}_T}{\sqrt{V[\sqrt{T} \bar{d}_T]}},$$

---

<sup>61</sup>Laurent et al. (2012) evaluate the robustness of two loss functions, the Frobenius distance and absolute error loss function, which is the sum of element-by-element forecast errors and is not consistent. A number of popular MGARCH models are applied to the cross-rates between the Euro, Yen and Pound. They also allow  $\Delta$  to vary. The authors find that in terms of the Frobenius distance, when  $\Delta < 60$ , the ranking amongst models is relatively unchanged. However, at lower sampling frequencies the rankings are altered significantly. The effect of an inconsistent loss function is manifested by the superior ranking of the RiskMetrics model under the absolute error loss function where  $\Delta > 120$ , since this is in sharp contrast to the same models "middle-of-the-pack" performance under the Frobenius distance.

where  $d_T = 1/T \sum_{t=1}^T d_t$  and the asymptotic variance of  $\sqrt{T}d_t$  can be calculated using a consistent estimator such as the Newey West estimator. Under  $H_o$ , the test statistic has a standard normal distribution asymptotically.

The Reality Check (RC) test of White (2003) may be considered advantageous compared to the DM test given that it tests for Superior Predictive Ability (SPA), which is generally of greater interest to the econometrician than EPA. In addition, the RC can compare more than two competing forecasts and is constructed in the following manner; a benchmark model  $h_t^0$  is first selected and the null hypothesis is that none of the alternative forecasts, numbered  $i = 1, \dots, l$ , outperform that benchmark in terms of a given loss function:

$$\begin{aligned} H_0 &: E[L(\hat{\sigma}_t^2, h_t^0)] \leq \min_{i \in \{a, b, c, \dots\}} E[L(\hat{\sigma}_t^2, h_t^i)] \\ H_1 &: E[L(\hat{\sigma}_t^2, h_t^0)] > \min_{i \in \{a, b, c, \dots\}} E[L(\hat{\sigma}_t^2, h_t^i)]. \end{aligned} \quad (71)$$

The performance of model  $i$  relative to the benchmark is again the difference in the loss functions:

$$d_{t,i} = L(\hat{\sigma}_t^2, h_t^0) - L(\hat{\sigma}_t^2, h_t^i). \quad (72)$$

The expected excess performance of  $i$  compared to the benchmark is then given as  $E[d_{t,i}] = u_i$ , while the average relative performance,  $\bar{d}_i$ , can be consistently estimated by the sample average  $\bar{d}_i = \frac{\sum_{t=1}^T d_{t,i}}{T}$ . This allows for the test statistic,  $T^{RC}$ , to be calculated as:

$$T^{RC} = \max_{i=1, \dots, l} (T^{1/2} \bar{d}_i). \quad (73)$$

After stacking the expected excess performance and average relative performance of each model relative to the benchmark into  $l \times 1$  vectors,  $\mathbf{u}$  and  $\bar{\mathbf{d}}$ , one can define the asymptotic  $l \times l$  covariance matrix as  $\mathbf{\Omega} = \text{avar}(T^{1/2}(\bar{\mathbf{d}} - \mathbf{u}))$ . The  $T^{RC}$  test then has an asymptotic null distribution given by  $T^{1/2}\bar{\mathbf{d}} \sim N_0(0, \hat{\mathbf{\Omega}})$ , where  $\hat{\mathbf{\Omega}}$  is a consistent estimator of  $\mathbf{\Omega}$ .

The problem with the RC test is that it is testing multiple inequalities and has a composite hypothesis whereby the asymptotic distribution of  $T^{RC}$  under  $H_0$  is not unique, (Hansen and Lunde (2005)). The distribution under the null assumes that  $u_i = 0$  for all  $i$ , even though it could be the case that some models significantly underperform relative to the benchmark, i.e.  $u_i < 0$ . Hansen (2005) suggests that this makes the testing procedure vulnerable to the inclusion of poor models, which affects the power of the test. This problem is addressed by using a sample-dependent null distribution which yields an upper and lower bound for the p-value. Hansen (2005) also notes that the p-values of the  $T^{RC}$  test are generally too large. Therefore he also derives a consistent p-value for the modified test. The null distribution,  $N_0(\hat{\mathbf{u}}^c, \hat{\mathbf{\Omega}})$ , thus has mean vector  $\hat{\mathbf{u}}^c$ , where each element is a function of  $\bar{\mathbf{d}}_i$  and an indicator function that is dependent on the value of the test statistic relative to some threshold<sup>62</sup>. In this way the sample determines the null distribution,

<sup>62</sup>More precisely:  $\hat{u}_i^c = \bar{d}_i 1\{T_n^{SPA} \leq \sqrt{2 \log \log n}\}$ .

where  $1\{\cdot\}$  is an indicator function.

unlike the RC test which assumes the same distribution. The test statistic is then given by:

$$T^{SPA} = \max \left[ \max_{i=1, \dots, l} \frac{T^{1/2} \bar{d}_i}{\text{var}(T^{1/2} \bar{d}_i)}, 0 \right], \quad (74)$$

which is distributed as  $N(\hat{\mathbf{u}}^c, \hat{\mathbf{\Omega}})$  under the null. The upper bound p-value makes the assumption of (71), and thus corresponds to the  $T^{RC}$  p-value and has the same asymptotic distribution. The lower bound p-value assumes that models that are worse performers than the benchmark, are poor models in the limit, which prevents them from affecting the asymptotic distribution of the test statistic. Given that we only have at our disposal one realization of the out-of-sample performance of each model, as measured by the loss functions, the  $T^{SPA}$  statistic relies on resampling methods to approximate the sampling distribution of this out-of-sample performance. More specifically, the asymptotic variance of the test statistic  $\text{var}(T^{1/2} \bar{d}_i)$ , is calculated using the stationary bootstrap of Politis and Romano (1994), although the block bootstrap can be used as well.

## 4.2 Economic evaluation of volatility forecasts

While variance forecasts are usually evaluated by statistical measures, the fact that the covariance matrix is a critical input in many economic and financial applications (such as mean-variance portfolio optimization and risk management) would imply that economic evaluation of the covariance matrix is intuitively appealing. Economic evaluation does require some auxiliary assumptions such as the type of investor utility function in the case of asset allocation. For risk management applications such as VaR, the type of choice of distribution will also be non-trivial.

Traditionally the out-of-sample performance of portfolios constructed using different forecasts have been evaluated within the mean-variance framework. While changes in the conditional mean do have a measurable effect on the optimal portfolio weights, Engle and Colacito (2006) show that a superior covariance forecast will lead to a smaller portfolio variance for any non-zero vector of the means. The problem of finding a global minimum-variance portfolio can be defined in the following equations:

$$\min_{w_t} w_t' \Sigma_t w_t. \quad (75)$$

$$\text{subject to } w_t' \iota = 1,$$

where  $\iota$  is a column of ones and  $w_t$  the vector of portfolio weights.

Patton and Sheppard (2009) show that if the portfolio weights are constructed using the actual covariance matrix  $\Sigma_t$ , then the variance of a portfolio using weights based on any other covariance forecast will be greater. The economic value of covariance forecasts derived from ARFIMA models has been assessed in Fleming et al. (2003), who sought to answer the question as to whether the realized covariance matrix can lead to superior asset allocation decisions. They consider an investor who allocates his wealth across four asset classes:

stocks, bonds, gold and cash. When using the realized covariance matrix and the non-parametric approach of Foster and Nelson (1994) to make forecasts of the conditional covariance matrix, the resulting tactical asset allocation decisions yield significant gains in terms of risk-adjusted returns, relative to returns on portfolios formed on the basis of a covariance matrix that was forecasted using the outer product of daily returns.

Investors however, do not typically hold currencies in isolation, but instead have currency exposure through their equity and bond holdings that are denominated in foreign currency. Examples of these types of models include the international asset pricing model (IAPM) of Solnik (1974), where investors with a globally diversified stock and bond portfolio (the world market portfolio) hold positions in forward contracts or bills denominated in foreign currency in order to hedge their currency exposure. Black (1990) makes additional assumptions in order to show that these hedges are also universal in that investors from any part of the world hold the same currency ‘portfolio’ as such and never fully hedge their currency exposure. However these results rely on extremely relaxed assumptions about the state of the world; no taxes, trading costs or restrictions on capital flows. A different notion is that of the cross-hedge, which is discussed in Anderson and Danthine (1981), where the holdings of stocks and bonds, as well as the international composition of the portfolio, are already determined ex-ante, and the hedges are chosen in order to improve the risk-return characteristics of this pre-determined portfolio. The vector of optimal weights for the hedge instruments  $\mathbf{f}$ , is given by the following expression:

$$\mathbf{f} = \lambda \Sigma_h^{-1} \boldsymbol{\mu}_h - \Sigma_h^{-1} \Sigma_{sh} \mathbf{s}, \quad (76)$$

where  $\lambda$  is the measure of risk aversion, while  $\Sigma_h$  and  $\boldsymbol{\mu}_h$  denote the covariance matrix and expected returns vector for the hedge instruments. The matrix  $\Sigma_{sh}$  holds the covariances between the assets in the underlying portfolio and the hedge instruments, and  $\mathbf{s}$  is the vector of the underlying asset positions. The first term can be viewed as a speculative component which disappears when expected returns are zero, while the second component is purely for hedging purposes. Glen and Jorion (1993) show that hedging a global portfolio (with or without diversification at the asset level) using currency forwards brings about a statistically significant increase in the in-sample Sharpe Ratio. However, the evidence for out-of-sample improvements was mixed and depended on the composition of the portfolio. In the event that there is only a single underlying asset and one hedge instrument, the hedge ratio can be obtained by regressing that asset on the hedge instrument. The case of a commodity owner looking to hedge his/her spot exposure using forward contracts is an example of this scenario, however, by using OLS one makes the assumption that the hedge ratio is constant, which is clearly at odds with the extensive literature concerning time varying variances and covariances. Based on this insight Baillie and Myers (1991) make use of bivariate GARCH (1,1) models<sup>63</sup> to obtain the conditional covariance matrix, which is used to calculate the optimal hedge ratio (OHR) as the ratio of the conditional covariance between the cash and futures market to the conditional variance of

<sup>63</sup>A diagonal vech model of Bollerslev et al. (1988) and the BEKK model defined later in Engle and Kroner (1995).

the futures instrument. A cash position in six different commodities<sup>64</sup> was hedged using a short position in the futures contract, where the hedge ratio was calculated using either OLS or the DVEC model. In an out-of-sample evaluation, the GARCH hedge brought about a significantly larger reduction in the portfolio variance when compared to the OLS hedge for every commodity except Gold. Kroner and Sultan (1993) conduct a similar study but focus on an investor with currency exposure, who wishes to hedge using futures contracts. The authors compare the in-sample and out-of-sample utility based on three hedging strategies; the naive hedge (hedge ratio=1), the conventional hedge (hedge ratio obtained using OLS) and a conditional hedge obtained from the conditional covariance matrix calculated using a bivariate GARCH model. On the basis of average utility, the conditional hedge outperformed the other hedging strategies, both in-sample and out-of-sample, even after accounting for transaction costs.

Given that our aim is to provide an economic evaluation of the conditional covariance matrix forecasts, our approach will be somewhat unique and could instead be considered a hybrid of the Glen and Jorion (1993), Fleming et al. (2003), Anderson and Danthine (1981) and Kroner and Sultan (1993) approaches. The main difference between our approach and the currency hedging in Kroner and Sultan (1993) is that the investor doesn't hold hard currency but instead carries currency exposure through his other assets. It may not be immediately obvious as to why our hypothetical investor would have currency exposure without actually holding assets that are denominated in a foreign currency, as in Solnik (1974) or Glen and Jorion (1993). To motivate our investor's desire to hedge against currency risk, we instead make use of the factor models that underpin the APT of Ross (1973), and view the USDZAR exchange rate returns as a macro factor that impacts asset returns. To the extent that individual firms pay and receive funds from creditors and debtors in different countries, each firm (and the returns on its shares) will be impacted by a currency depreciation/appreciation in a different way. At the macro level a currency depreciation is generally seen as favorable to overall output in the medium term, provided the country is a net exporter. This is why countries have been known to depreciate their currencies in order to boost output and employment in the near future. When using factor models the returns of a particular share are regressed on a set of factors in order to obtain the beta or factor-loading for each of those factors. The remaining variance that is not explained by the factors,  $\sigma_e^2$ , is then the remaining firm-specific or idiosyncratic risk. Here we proceed differently; we estimate the exchange rate factor-loading for the broader South African equities market by regressing the excess monthly<sup>65,66</sup> All Share Index (ALSI) return on the monthly USDZAR forward<sup>67</sup> returns. The remaining variance is not firm-specific risk as before, but rather the systematic risk from the other factors

<sup>64</sup>Beef, Coffee, Corn, Cotton, Gold and Soybeans.

<sup>65</sup>The choice of monthly returns is obviously guided by the availability of the one-month forward contract as the hedging instrument. It is also in line with much of the currency hedging literature; see Glen and Jorion (1993) and Campbell et al. (2007).

<sup>66</sup>Excess returns are those over the risk-free rate, which is the 3-month RSA T-bill.

<sup>67</sup>Usually currency returns must incorporate any interest rate differential between countries, but this differential is already included in the pricing of the forward contract.

not included in our regression. The actual estimates<sup>68</sup> obtained from such a regression are given by<sup>69</sup>:

$$r_{ALSI,t} - r_{f,t} = \frac{0.003}{(0.0043)} - \frac{0.244}{(0.0902)} r_{USDZAR,t} + e_t. \quad (77)$$

Clearly there is a contemporaneous inverse relationship between the USDZAR rate and the ALSI<sup>70</sup>. Looking at Equation (77), it is immediately obvious that an investor with diversified South African equity holdings can reduce the variance of his/her portfolio by purchasing USDZAR forwards, given that  $var(r_{ALSI,t} - r_{f,t}) = \beta^2 \sigma_{USDZAR}^2 + \sigma_e^2 > \sigma_e^2 = var(r_{ALSI,t} - r_{f,t} + 0.244 r_{USDZAR,t})$ . The absence of other systematic risk factors such as the output gap or unanticipated inflation in (77) may lead to a case of omitted variable bias for the exchange rate beta. However, the only condition needed to validate the risk-reducing effect of buying USDZAR forwards is that  $\beta \neq 0$ .<sup>71</sup>

The formal framework within which we conduct this hedging exercise and economic evaluation of the covariance matrix forecasts, rests upon the following assumptions:

1. Our hypothetical investor will have two possible pre-determined portfolios at time  $t$ ; a locally-diversified<sup>72</sup> stock portfolio or a portfolio consisting of 60% diversified stock holdings and 40% diversified bond holdings.
2. The investor cross-hedges the exposure of his portfolio to fluctuations in the USDZAR exchange rate by going long  $b$  units of USDZAR forwards. Thus the return on the hedged portfolio is given by:

$$r_{h,t+1} = r_{t+1} + b_t f_{t+1}, \quad (78)$$

where  $r$  is the return on the unhedged portfolio and  $f$  the return on the forward contract.

3. The investor is risk-averse and must be induced into holding riskier assets with higher expected returns. Therefore, he/she faces the mean-variance or quadratic utility function:

$$EU(r_h) = E(r_h) - \gamma Var(r_h),$$

<sup>68</sup>The sample begins in January 1997 and ends in February 2013.

<sup>69</sup>Standard errors are given in parentheses.

<sup>70</sup>Campbell et al. (2007) produce a similar finding for the Euro and Swiss franc and find those currencies are negatively correlated with their domestic equity returns. Inflation, and its effects on asset prices, as well as currencies, is one possible explanation for this result. If one views equities as real assets, then an increase in inflation will erode their value and the value of the currency through the PPP mechanism. We posit that the search for yield by foreign investors and the associated portfolio flows also play their part. When South African equities or bonds are seen as attractive investments by foreign investors the associated inflow of foreign capital leads to an appreciation of the rand (USDZAR falls) and a rise in the equity and bond indices. When foreign investors sell their equity and bond holdings and exchange their rands for dollars, the rand will depreciate (USDZAR rises) and the equity indices fall.

<sup>71</sup>The exchange rate has been found to be a significant factor in empirical studies, see Chen et al. (1986).

<sup>72</sup>By locally-diversified we mean a portfolio made up of South African assets sufficient in number such that there is no remaining idiosyncratic or firm-level risk.

where  $\gamma$  is the degree of risk aversion ( $\gamma > 0$ ). The investor therefore solves the following utility maximization problem by choosing  $b$ , where,

$$\max_b EU(r_h) = \max_b \{E(r) + bE(f) - \gamma [\sigma_r^2 + b^2\sigma_f^2 + 2b\sigma_{rf}]\}.$$

Thus the optimal hedge ratio is given by:

$$b^* = \frac{E(f_{t+1}) - 2\gamma\sigma_{rf}}{2\gamma\sigma_f^2},$$

which reduces to  $b^* = -\frac{\sigma_{rf}}{\sigma_f^2}$ , if expected returns for the forward contract are zero.

4. The investor calculates the optimal hedge ratio at time  $t$  based on the conditional covariance matrix from a number of popular MGARCH models, as well as from a ARFIMA model for the vectorized covariance matrix:

$$b_t^* = -\frac{\sigma_t(r_{t+1}, f_{t+1})}{\sigma_t^2(f_{t+1})} \quad (79)$$

The forward premium puzzle and the profitability of the carry trade<sup>73</sup> suggests that interest rate differentials between South Africa and typical "funding" countries/currencies like Japan or Switzerland could yield feasible ex-ante predictions of the USDZAR exchange rate. However, as our focus is solely on the second moment of portfolio returns, and we abstract from predictions of the USDZAR returns, our investor's problem is one purely of risk-minimization and has no speculative component. We make no mention of wealth given that unlike the portfolio optimization problem described in (75) our investor has had the selection of  $w$  made for him, and his utility is solely a function of the first and second moment of his portfolio returns. The absence of a time script indicates that the portfolio weights are fixed throughout the entire sample and implicit in this assumption is a monthly rebalancing of the portfolio back to the pre-determined weights. The question of how many forwards to buy is then a completely separate issue from that of the asset allocation decision<sup>74</sup>.

## 5 Date and Results

### 5.1 Univariate Data

The intraday data used in this study were obtained from disktraders.com<sup>75</sup> and consists of 5 minute closing prices that span the period 2/01/2008 to 31/12/2012<sup>76</sup>. The efficient prices are proxied by the midpoint

<sup>73</sup>See Hansen and Hodrick (1980), Engel (1996), Burnside et al. (2006), Galati et al. (2007) and Hassan and Smith (2011).

<sup>74</sup>We concede that whilst the assumptions made here are not entirely realistic, they should make our analysis tractable. In addition they should facilitate a meaningful comparison amongst the unhedged portfolio and the assortment of MGARCH hedged portfolios.

<sup>75</sup>Other studies to make use of the same data vendor include Koopman et al. (2005) and Marcucci (2006).

<sup>76</sup>All model estimation (excluding the HEAVY model, DCC model) was done using Oxmetric's G@RCH or PC-GIVE modules. For more on the Ox programming language and software see Doornik (2009). The HEAVY



between the last ask and bid quotes recorded before the relevant time mark. Formally, this is known as the previous tick method<sup>77</sup> which differs from the linear interpolation method, (Dacorogna and Gencay (2001)). The selection of a five minute frequency is consistent with other analyses of realized volatility given in Andersen et al. (2001) and Andersen et al. (2001). The FX market runs 24 hours a day from 22:30 GMT on a Sunday when trading commences in Asia and ends at 22:30 GMT on a Friday, when trading wraps up on the West Coast of the United States, (Guillaume et al. (1997)). This particular data was originally stamped according to New York or Eastern Standard Time (EST) and therefore runs from 17:25 EST on Sunday to 17:25 EST on Friday, the result being that there is a Sunday trading session consisting of just 6 and a half hours of trading. Given that the HEAVY model as well as the ARFIMA model will rely on the Friday period's realized volatility to make a prediction for Monday's volatility, the time was adjusted to GMT+1 time in order to reduce the amount of trading that happens on a Sunday to just 30 minutes, whilst also ensuring that there is no Saturday trading period. This adjustment also makes the sample more congruent with the exchange rate in question; the South African Rand price of US dollars (or USDZAR)<sup>78</sup>. Given that high-frequency prices are only available when the market is open any realized variance measures will represent the variance of open-to-close returns. Our returns are of the close-to-close variety and accordingly we need to calculate the close-to-close volatility. Fortunately, in the case of the 24-hour FX market, this corresponds extremely closely to the open-to-close variation given that the gap between the closing price on day  $t$  and day  $t + 1$  will be negligible. For consistency the same sample was used to obtain both the 5 minute and daily returns, which were calculated as the change in the log price.

We make the explicit assumption that the prices used to calculate returns contain microstructure noise and that this noise is i.i.d as well as being independent of the efficient price. In order to calculate the realized volatility we have selected the Hansen et al. (2008) MA(1) filter, since it provides an unbiased estimate of the realized volatility in the presence of microstructure noise, under the aforementioned assumptions. The HEAVY model and the ARFIMA model rely on the realized volatility as their model inputs, but this same measure will also be used for forecast evaluation. Therefore, if the realized volatility has a marked upward bias then any model that utilizes realized volatility to make forecasts will have a built-in advantage relative to a standard GARCH model, which uses an unbiased measure of the volatility, namely the returns squared.

Table 1 presents the unconditional moments as well as some additional statistics for the raw 5 minute returns. To economize on space only the features of this intraday data that are relevant to our analysis and methodology are discussed. Consistent with the presence of microstructure noise we estimate the first-degree autocorrelation in the 5 minute returns to be -0.1431, which is significantly higher than the -0.04 found in and DCC model were estimated using MATLAB and the MFE Toolbox GARCH routines of Kevin Sheppard, see [http://www.kevinshppard.com/wiki/MFE\\_Toolbox](http://www.kevinshppard.com/wiki/MFE_Toolbox). MLE was used for all models except the GO-GARCH model where Non-linear Least Squares estimation was utilized, see Boswijk and Van der Weide (2006). All other computations including the calculation of the MA(1) realized variance and monthly realized covariance matrix, the hedge ratios, utility, univariate and multivariate loss functions as well as the SPA test were computed in MATLAB.

<sup>77</sup>Hansen and Lunde (2004) suggest using the previous tick method in the context of kernel-based realized volatility.

<sup>78</sup>These short Sunday periods have been omitted from the final sample used to estimate and evaluate the models.

Andersen and Bollerslev (1997) for the 5 minute DMUSD exchange rate returns. This is not surprising given that prior to the advent of the Euro the DMUSD was the most liquid currency pair and the USDZAR market doesn't possess anywhere near the same amount of liquidity. The intraday returns are highly non-normal with a large degree of kurtosis, higher than that reported in Andersen and Bollerslev (1997). Both the LM test for ARCH effects and the Q-stat calculated on the squared returns strongly suggest the presence of heteroskedasticity in the intraday returns, which implies that the MA(1) model of Hansen et al. (2008) used to calculate realized volatility is misspecified. However, as was discussed in Section Three this should not affect the statistical performance of the estimator too adversely, in that it will still maintain its unbiasedness for the integrated variance. The sample ACF of the 5 minute returns in Figure 1 is clearly indicative of an MA(1) process. When the returns process is modelled as an ARMA(0,1) process, the MA(1) coefficient estimate is -0.1469 and has a highly significant t-statistic of -246.82.

Daily returns were constructed using the last 5 minute closing price of the current day and the last 5 minute closing price of the prior day. Table 2 presents the descriptive statistics and Figure 2 the density plot of the daily returns<sup>79</sup>. The final sample to which the summary statistics refer amounted to 1278 daily observations. From Table 2 we can see that returns are again clearly non-normal with a significant amount of skewness and kurtosis. This is largely due to a number of massive outliers that took place during the Global Financial Crisis in 2008<sup>80</sup>. This time however, the first-order autocorrelation (as well as all the autocorrelations for lags 2 through 20) are insignificant, thus no ARMA modelling of the conditional mean will be necessary. The LM test and Q-stat again point overwhelmingly towards time-varying volatility in the returns process.

As suggested by Hansen et al. (2008), in order to calculate the realized volatility for any given day, an MA(1) coefficient should be estimated using only the returns during that period, thus the MA-based realized variance is given by<sup>81</sup>:

$$RV_{MA,t} = (1 - \hat{\theta}_t)^2 \sum_{j=1}^m \hat{e}_j^2, \quad (80)$$

where  $\hat{e}_j^2$  is the squared MA(1) filtered residual and  $\hat{\theta}_t$  the MA(1) coefficient for trading day  $t$ . Descriptive statistics for the MA(1) realized variance are contained in Table 3<sup>82</sup>. Again the data are highly non-normal

<sup>79</sup>Note that a handful of days with highly suspect price patterns involving numerous outliers or days in which identical prices were recorded throughout the day were also dropped from the sample, these were the 14/01/2008, 26/05/2008, 22/04/2011, 9/04/2008, 1/05/2008, 21/02/2011, 04/07/2011. The following traditional holiday periods were also excluded; 24th, 25th, 26th and 31st of December as well as the 1st and 2nd of January.

<sup>80</sup>On the 15th of October 2008 the Rand lost more than 16% of its value against the Dollar in a single day. In fact that particular month is by far the most volatile with 3 of the top 5 biggest depreciations in the sample and 4 of the top 5 biggest appreciations all occurring in the same month.

<sup>81</sup>Over such a large sample it's entirely possible that the MA coefficient on a given day may not be statistically significant. However we make the MA filter correction for every trading day in the sample, regardless of statistical significance.

<sup>82</sup>The sheer magnitude of the volatility experienced during 2008 is evident when looking at the progression of annualized realized volatility from 2008-2012 which is 9.109%, 5.167%, 2.571%, 3.414% and 2.143% while the overall average is 4.463%. 9

with significantly more skewness and kurtosis than intraday returns and daily returns. The first-order autocorrelation is significant, as are the following 99 sample autocorrelations. This is illustrated in Figure 3, which indicates that realized volatility follows a long memory process. This finding is directly in line with those of Andersen et al. (2001) and Andersen et al. (2001).

We turn next to the properties of the log of the MA(1) realized variance which we denote as  $\log(RV)$ , see Table 4. The transformed series still possesses a moderate amount of skewness, however, the kurtosis is near normal. Despite the fact that the data is near-normal, the test statistic nevertheless suggests that we should reject the null of normally distributed data at the 5% level. Compared to the findings of Andersen et al. (2001) for the DMUSD we find more skewness and more kurtosis<sup>83</sup> (0.8785 vs 0.348 and 4.62 vs 3.26). Figure 4 displays evidence of long memory in the  $\log(RV)$  process, with the magnitude of the autocorrelation at lags 1 through 100 greater than that of the ACF for the realized variance. In addition, the decline in the  $\log(RV)$  ACF is noticeably slower than that for realized volatility, suggesting that there is possibly a higher degree of fractional integration or longer memory in the transformed series.

Having obtained the MA(1) realized standard deviation, we then calculated the series of standardized daily returns;  $z_t = \frac{r_t - \mu}{RV_{MA,t}^{1/2}}$ , whose distributional characteristics are given in Table 5. If our estimates of the realized volatility are unbiased we would expect the mean and standard deviation to be approximately zero and one respectively. Table 5 confirms that this is indeed the case with a mean of -0.0334 and a standard deviation of 0.9135, suggesting that by using the MA(1) filter we have eliminated most of the upward bias in the realized variance caused by the microstructure noise. The skewness and kurtosis are very close to Gaussian proportions and we narrowly miss out on accepting the null of a normally distributed series.

Figure 5 presents the quantile-quantile (QQ) plots for daily returns, realized variance, daily returns standardized by the realized standard deviation, as well as the log of the MA(1) realized variance. The close approximation of the standardized returns to that of the standard normal distribution is clear from the bottom right panel, given that the QQ plot follows a 45 degree line while the S-shape of the daily returns and realized volatility QQ plots further highlight the excess skewness and kurtosis relative to the standard normal distribution. The flatness of the trend line in the plots for daily returns and the MA(1) realized variance further illustrates the magnitude of the outliers within these variables.

To summarize, our findings for the USDZAR exchange rate broadly support those of Andersen et al. (2001) and Andersen et al. (2001), in that daily returns standardized by the realized standard deviation are approximately normal as is the log of the daily realized volatility, with the former being a far closer approximation than the latter. Additionally, there is strong evidence for a long memory process in the realized variance as well as its logarithm.

---

of the top 10 most volatile trading days occur during October 2008.

<sup>83</sup> Andersen et al. (2001) examine the log of the realized standard deviation. However this has the same kurtosis and skewness as the log of realized variance.

## 5.2 Univariate Estimation

Andersen et al. (2003) have shown that reduced-form ARFIMA models of realized volatility provide superior out-of-sample forecasts for the volatility in exchange rate data. Models of this type, which we henceforth denote as ARFIMA-RV, can be said to have two distinct advantages over the traditional GARCH methodology of Bollerslev (1986). Firstly, the usage of a far less noisy proxy for the previous period's volatility, i.e. realized volatility versus daily returns (or residuals) squared. Secondly, it allows for a relatively straightforward way of explicitly modelling the long memory behavior typical of the volatility process. There exists within the vast ARCH/GARCH family models that can make use realized volatility, the HEAVY model of Shephard and Sheppard (2010) being one such example. There are also models that treat volatility as a fractionally integrated process such as the FIGARCH model of Baillie et al. (1996). However, there is no single one model that possesses both these characteristics. By comparing the HEAVY and FIGARCH forecasts to that of the ARFIMA-RV we hope to gain some insight into the main source of the ARFIMA-RV's forecasting edge that has been reported in other studies. (i.e. is it due to the higher quality volatility proxy or the long memory aspect?). The four models we will estimate here; GARCH, FIGARCH, HEAVY and ARFIMA-RV, can therefore be dichotomized neatly, as is depicted in Figure 6.

In terms of methodology, we will first select a representative model from each family; GARCH( $p, o, q$ ), FIGARCH( $p, d, q$ ), HEAVY( $p, q$ ) and ARFIMA-RV( $p, d, q$ ) based on in-sample fit and then compare the out-of-sample forecasting performance of each representative model based on statistical measures covered in Section Four. For each of the models the order of the respective lags were selected over the entire sample period. We make use of information criteria for the purpose of model selection, starting with higher orders of  $p, o$  and  $q$  and decreasing the number of lags on the basis of the Schwarz or Bayesian Information Criterion (SBIC)<sup>84</sup>.

Table 6 provides the MLE parameter estimates of our GARCH model, for each of the following error distributions-Normal, Student's-t and the GED distribution. Higher order lags for  $p$  and  $q$  were considered but were unable to provide a smaller SBIC<sup>85</sup>. An interesting finding here is that of a statistically significant asymmetric innovation term, which is not often reported on when analyzing the volatility of currencies. The sign of the coefficient asserts that a negative return ( $e_t < 0$ ) in period  $t$  is likely to be followed by lower volatility during period  $t + 1$ , in other words following an appreciation of the rand against the dollar, the following period is more like to have lower volatility than if there had been a depreciation. For the moment we reserve discussion on this result. Given this finding, we will henceforth refer to this particular model as the GJR-GARCH model. The GJR-GARCH(1,1,1) model with Student's-t distributed errors has the lowest SBIC, with all the individual parameter estimates being significant and thus will be considered the benchmark GARCH model. It's interesting to note that all three cases possess a near unitary persistence parameter, with persistence of 0.973584 in the case of the Student's-t distribution giving some insight into

<sup>84</sup>This criterion was chosen over the Akaike Information Criterion (AIC) as it values parsimony more highly than the AIC by penalizing additional parameters more harshly.

<sup>85</sup>The parsimonious GARCH(1,1) model has often proven superior to more complex formulations, see Hansen and Lunde (2005).

intertemporal dependency of the USDZAR’s conditional variance.

Table 7 presents the FIGARCH(1,d,1) MLE estimates under each different error distribution<sup>86</sup>. The SBIC for the model with Student’s-t distributed errors is again the best based on in-sample fit. Of particular interest is the estimate of 0.3214 for the long memory parameter  $d$  which unlike the other individual parameter estimates (excluding the shape parameter) is statistically significant. This is considerably less than Baillie et al. (1996)’s estimate of 0.652 for the DMUSD rate.

Next, we find the best in-sample HEAVY model based on the SBIC, but unlike the GJR-GARCH and FIGARCH model, we only consider Gaussian errors (due to software limitations). The parsimonious HEAVY(1,1) was selected over specifications with higher order lags and the MLE parameter estimates are to be found in Table 8. As was the case for the GJR-GARCH model, the persistence is again very close to one, however, the persistence is evenly split between the AR and MA coefficient. This is markedly different from the traditional GARCH methodology, where  $\hat{\alpha}$  and  $\hat{\beta}$  are usually close to zero and one respectively.

Finally, we consider the direct modelling of realized volatility itself by way of an appropriate ARFIMA model. In this regard we follow Andersen et al. (2003), by modelling the natural logarithm of realized volatility, however, our approach does differ somewhat in that we are making use of the MA-based realized variance of Hansen et al. (2008), and not the traditional sum-of-squares estimator of Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002)<sup>87</sup>. The MLE estimates for the benchmark ARFIMA(2,d,2) model are given in Table 9, where the fractional differencing parameter estimate is 0.4909. Note that the presence of the ARMA parameters would imply that in order to facilitate a direct comparison between the USDZAR and the currencies examined in Andersen et al. (2003), an additional ARFIMA estimation was performed, one that did not include any ARMA components<sup>88</sup>. These results suggested that our mostly unchanged estimate of 0.4888 remains higher than the 0.387 and 0.413 found for the DMUSD and JPYUSD in Andersen et al. (2003), suggesting greater persistence in the log of realized USDZAR volatility, relative to the more liquid currency pairs. For  $d > 0.5$  the volatility process is no longer fractionally integrated or covariance stationary<sup>89</sup>, and with a standard error of 0.01233, we cannot reject a null hypothesis of  $d > 0.5$  on the basis of a simple t-test. Given that the standard GARCH specification corresponds to the case where  $d = 0$ , and the IGARCH model refers to the situation where  $d = 1$ , then an estimate of approximately 0.5 suggests the dynamic evolution of the USDZAR exchange rate variance falls somewhere in the center of the ‘persistence-spectrum’ spanned by these two extreme cases. Put another way, the empirical ACF of the USDZAR realized volatility displays a rate of decay slower than the exponential decay of the GARCH model, but not as slow as the unit-root IGARCH model, which implies complete persistence<sup>90</sup>.

<sup>86</sup>Estimation of the FIGARCH model introduces a computational issue in that the infinite expansion of  $(1 - L)^d$  must be truncated. Here we follow established practice and restrict the infinite series to a length of 1000 observations.

<sup>87</sup>Additionally, Andersen et al. (2003) use 30 minute returns which are not as affected by microstructure noise than 5 minute returns, hence the need for the MA-based estimator.

<sup>88</sup>It must be noted that Andersen et al. (2003) calculate  $d$  using the GPH estimator of Geweke and Porter-Hudak (1983).

<sup>89</sup>See Baillie (1996).

<sup>90</sup>The original measure of long-memory, the Hurst exponent was introduced by Hurst (1950) where  $H = d + 0.5$ . If  $0 < H < 0.5$

For completeness Table 10 presents a number of diagnostic checks on the standardized residuals from each of the four models estimated. All four models have sufficiently accounted for the heteroskedasticity in-sample, as evidenced by the lack of remaining time-varying volatility in the standardized residuals.

### 5.3 Univariate Forecasts

All univariate models will be estimated on a rolling sample of 1022 daily observations. Given that the forecast period extends from the first trading day of 2012 to the last trading day of that same year, on the first iteration the estimation sample will end with the 30th of December 2011 and extend back a further 1021 observations to the beginning of the sample. On the second iteration the first trading day in the prior estimation sample will fall away and the first trading day of 2012, the 3rd of January will enter into the estimation sample. This process will be repeated 256 times until we have created our series of 1-step-ahead out-of-sample forecasts.

The out-of-sample period has been split into four sub-periods of equal length and the forecasts have been plotted with the MA(1)-based realized USDZAR variance in Figure 11. During the first quarter, the realized volatility mostly vacillates between high and low values, seemingly with little persistence although the volatility during the first half of this particular period is on average higher than the second half. Throughout the rest of the year there are three periods of increased volatility; June, September and October. It's only during these periods that the differences between forecasts become noticeable. It's also quite evident from these figures that the pairwise forecasts from the HEAVY and ARFIMA-RV models, as well as the GJR-GARCH and FIGARCH models, are highly correlated with one another, which is obviously a function of the respective volatility proxies for these two pairs.

#### 5.3.1 Statistical Evaluation of Univariate forecasts

Before turning to the evaluation of the respective univariate forecasts in a relative sense, we conduct MZ regressions to gauge the success of the forecasts in an absolute sense. We also conduct "Encompassing Tests"<sup>91</sup> that will tell us whether the GJR-GARCH, FIGARCH or HEAVY model are still able to provide any incremental information that can improve upon forecasts made by the ARFIMA-RV model. In the main, our results (Table 11) are remarkably similar to those of Andersen et al. (2003). For the MZ regression, the only model for which we cannot reject the joint hypothesis that  $\beta_0 = 0$  and  $\beta_1 = 1$  at the 5% level is the ARFIMA-RV model. In terms of overall fit, the ARFIMA-RV model has the highest  $R^2$  ( $\approx 0.28$ ), although the HEAVY model's  $R^2$  of 0.26 is not far behind<sup>92</sup>. This finding, along with the fact that the  $R^2$  of the GJR-GARCH model is in fact higher than the long memory GARCH model would suggest that superiority of

---

then the series is anti-persistent or mean-reverting but if  $0.5 < H < 1$  then the series has long-memory, thus an estimate of 0.4909 for  $d$  implies that our estimate for  $H$  is near the upper end of this band.

<sup>91</sup>See Patton and Sheppard (2009).

<sup>92</sup>Andersen et al. (2003) evaluate three currencies and in terms of overall fit our results are similar to the  $R^2$  they obtained for the DMUSD rate ( $\approx 0.25$ ). The  $R^2$  for the JPYUSD and JPYDM was roughly 0.35.

the ARFIMA-RV model is mostly a function of the volatility proxy and not its long memory structure. Like Andersen et al. (2003) the encompassing tests (Second Panel) reveal that none of the alternative models to the ARFIMA-RV are able to add any statistically significant information to its forecasts and the  $R^2$  of those regressions are virtually identical to that of the MZ regression. In the case of the GJR-GARCH model this is somewhat surprising as it includes an asymmetric volatility response parameter, which the ARFIMA-RV model does not have.

The overall fit of the HEAVY and ARFIMA-RV models versus the GJR-GARCH and FIGARCH is vastly superior and may lend credence to the theory that these models have an unfair advantage because the variable used for ex-post evaluation is their model input, so any bias, no matter how small may influence the MZ regression results. As a robustness check we perform another set of MZ regressions whereby the ex-post variance or left-hand-side variable is no longer the MA(1) realized variance, but the daily return squared, Table 12 contains these results. Overall, the forecasting performance is dismal due to the noise in the forecast benchmark and we reject the null hypothesis at the 5% level for each of the models. However, the ordering of the models based on the  $R^2$  is mostly the same, where the HEAVY and ARFIMA-RV models have markedly higher in-sample fit relative to the FIGARCH and in particular the GJR-GARCH model. The ARFIMA-RV model is again the standout on the basis of  $R^2$  and none of the competing models can provide any incremental information.

Table 13 presents the SPA test p-values based on the two robust loss functions described in Section Four; the MSE and QLIKE<sup>93</sup>. These results corroborate the findings of the MZ regressions. There are four models in total, one of which must be designated as the benchmark model. The approach we follow here is to make each model the benchmark once and leave the remaining models as the alternatives. Note that all results will be interpreted based on the Consistent p-value of Hansen (2005). The results are striking; we can reject the null of no outperformance at the 5% significance level for all models except the ARFIMA-RV which has a p-value of 1. This provides strong evidence that on the basis of statistical measures, the direct modelling of realized volatility is superior to the traditional GARCH models that use daily returns squared as their volatility proxy. The fact that the ARFIMA-RV outperforms the HEAVY model, which uses the same volatility proxy suggests that the ARFIMA-RV model's long memory structure is the source of its edge over the HEAVY model.

---

<sup>93</sup>The implementation of the SPA test and the bootstrapping procedure needed to obtain an estimate of the asymptotic variance of the vector of outperformance requires that two practical choices be made; in terms of the number of bootstrap replications, and the block length. In the first instance we follow the suggested procedure in Hansen (2005) by increasing the number of replications until the p-values are invariant to further incremental increases. This happened at 5000 replications. In terms of the average block length, we use the procedure first outlined in Politis and White (2004) and in Patton et al. (2009), which chooses the block length in order to minimize the MSE. The lack of autocorrelation in adjacent values of the MSE losses led to a block size of just one observation, while the high degree of intertemporal dependence in QLIKE losses resulted in an optimized block length of 20 observations.

## 5.4 Multivariate Data

The template for modelling the USDZAR's comovements with other financial variables will be similar to that followed in the univariate setting. Using traditional MGARCH models, as well as an ARFIMA model for each unique element of the vectorized realized covariance matrix, we will examine whether direct modelling is superior to the MGARCH approach in a pure forecasting sense. In line with our assumptions in Section Four we select well-diversified equity and bond indices as a proxy for our investors holdings, as well as the 1-month USDZAR forward, which is used as the hedging instrument. The daily data were obtained from Thomson-Reuters and consists of 4219 closing prices for the 1-month USDZAR forward<sup>94</sup>, the South African All-Bond Index (ALBI) and the JSE All-Share Index (ALSI) over the period 31/12/1996 to 1/03/2013. Figures 12 and 13 present the log closing prices and monthly returns for all three series as well as the monthly returns for the hypothetical 60/40 equity and bond portfolio which we henceforth refer to as the 'Mixed Portfolio'<sup>95</sup>. Table 14 presents some statistics for the daily and monthly returns for ALSI, ALBI, USDZAR and Mixed Portfolio.

The impact of three episodes in the domestic and international markets are clearly visible in these figures. In 1998 and 2001 there were two currency crises which not only resulted in a massive depreciation of the rand (particularly in 2001) but spilled over into the equity and bond markets, although the spillover during the second crisis wasn't as severe as it had been 3 years earlier despite the depreciation being of a far greater magnitude. Table 15 presents the correlation between each index, the Mixed Portfolio and the USDZAR forward over the entire sample, while Table 16 presents the correlation for 2008-2009. Firstly, the non-unitary correlation between the ALSI and ALBI indicate potential gains from diversification at the asset level. Interestingly, the entire sample correlation between the ALSI and ALBI is positive but the two-year correlation for 2008-2009 is negative. This broadly supports the theory that investors were shifting out of equities and into less risky assets during this period of heightened risk aversion. Outside of this two year window, asset prices in the bond and equity markets largely moved in lockstep. The negative correlation between the USDZAR and the ALBI is greater than the correlation between the USDZAR and ALSI and relatively unchanged between the entire sample and sub-sample, however, the negative correlation between the USDZAR and ALSI more than doubles in magnitude during the 2008-2009 period. This is seemingly consistent with the theory that portfolio outflows from South African equities in particular were leading to a weaker rand. Admittedly the evidence presented here is anecdotal, however, for the risk-minimizing investor the gains from hedging do not rest on pinpointing the exact causality between the currency markets and the asset markets. The impact of these three macro events is also evident from the monthly returns series in Figure 13.

As was the case with the USDZAR daily returns the daily ALSI, ALBI and Mixed Portfolio data are

---

<sup>94</sup>For a thorough description of the calculation of the forward exchange rate visit <http://www.wmcompany.com/pdfs/WMReutersMethodology.pdf>

<sup>95</sup>Descriptive statistics for daily rand returns, albeit over a smaller sample, can be found in Table 2.



highly non-normal with strong statistical evidence for heteroskedasticity in all three time series as well as statistically significant first-order autocorrelation. In terms of the volatility calculation, unlike the univariate case where one would sum 288 squared intraday returns, the monthly realized variance will require at most the summation of 22 daily returns where the magnitude of the first-order autocorrelation is smaller compared to the 5 minute returns, therefore we don't make any adjustments for this autocorrelation as any bias will be negligible. Accordingly we calculate the realized variance using the sum-of-squared returns approach of Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002).

Monthly returns for the USDZAR, the individual components that make up the Mixed Portfolio (ALSI and ALBI) as well as the Mixed Portfolio itself are all non-normal but unlike the daily returns there is no significant first-order autocorrelation. Interestingly, the LM test and Ljung-Box Q-stat for the squared monthly returns don't provide evidence of heteroskedasticity<sup>96</sup>, however this is likely due to the small sample size, rather than a genuine lack of time-varying volatility, as can be seen from a casual inspection of Figure 13. The addition of bonds to the investor's equity holdings and its desirable impact on volatility are evident from Table 14 where the standard deviation of monthly returns decreases from 0.0599 to 0.0406. However the investor does sacrifice a higher return for this variance reduction; a mean return of 0.0097 is obtained for the pure equity case versus a mean return of 0.0066 for the Mixed Portfolio. Asset diversification also has an effect on the third and fourth moments of the Mixed Portfolio's return distribution. The probability of large positive or negative returns increases, as evidenced by the higher kurtosis in the Mixed Portfolio. The Mixed Portfolio is also more skewed to the left than the ALSI.

Table 17 provides summary statistics for the realized variance and log realized variance ( $\log(RV)$ ) for the USDZAR, ALSI and Mixed Portfolio. The general features of the data are remarkably consistent across all three series; where the realized variance is highly right-skewed and leptokurtic but the  $\log(RV)$  is approximately normal. Both the realized variance and  $\log(RV)$  possess a high degree of first order autocorrelation. Tables 3 and 4 and Table 17 provide an interesting parallel between daily realized USDZAR variance and monthly realized USDZAR variance; the realized variance at both frequencies are right-skewed and leptokurtic while the log of monthly realized variance is an even closer approximation to the normal distribution than was the case for the log of daily realized variance despite the fact that we were unable to accept the null hypothesis of normally distributed data. The effect of diversification at the asset level is again highlighted by the lower realized variance for the Mixed Portfolio over the ALSI. The key difference between the daily and monthly realized volatility is that the lower frequency data no longer possesses long memory, as can be clearly seen from examining Figures 3 and 14, the ACF's of the daily realized variance and monthly realized variance. The rapid exponential decay in Figure 14 suggests an AR(p) process for monthly realized volatility without a great deal of persistence.

Table 18 shows that monthly USDZAR, ALSI and Mixed Portfolio returns standardized by their respec-

---

<sup>96</sup>Except the Q-stat for the squared ALBI returns.

tive realized standard deviations are normally distributed based on the JB test statistic.

In order to calculate the realized covariances between each of our possible pre-determined portfolios and the USDZAR forward we first construct a historical daily return series for each unhedged portfolio by varying the weights for the ALSI and ALBI holdings, denoted  $(w_{ALSI}, w_{ALBI})$  between  $[1,0]$ , and  $[0.6,0.4]$ <sup>97</sup>. The monthly realized covariances were then calculated as the sum of the daily cross products between the ALSI or Mixed Portfolio returns and the USDZAR forward returns. The distributional characteristics and additional statistics for the covariances under each possible weighting scheme are given in Table 19 . Both data series are highly non-normal and have significant first-order autocorrelation. What is surprising is the finding of long memory for both covariances, which is depicted in Figure 15. The covariance between the ALSI and the USDZAR rate and between the Portfolio and the USDZAR rate both show a high degree of persistence with MLE estimates for the differencing parameter and standard errors given by 0.4281 (0.0541) and 0.4576 (0.0439) respectively.

## 5.5 Multivariate Estimation

Our next task is to select a representative model from each of the BEKK, DCC, RiskMetrics and GO-GARCH MGARCH families. Instead of choosing one specific model and then applying it to each of our portfolios we thought it appropriate to select a model from each family, for each possible bivariate returns series. For example, the DCC model requires that one specify a univariate structure for each variable, and for the ALSI case it is likely that there will be a statistically significant asymmetric volatility response parameter, however, this may not hold true for the Mixed Portfolio. As it turns out, the same model structure as well as the number of lags for each category of model was the same for both portfolios, based on the SBIC. In the univariate section another dimension was added to the model selection by way of the error distribution. Given that the Student's-t distribution dominated both the Gaussian and GED, we will make use of the Student's t-distribution where possible. As in the univariate section we will choose the best model based on in-sample fit over the entire sample.

Beginning with the BEKK estimates in Table 20, we can see that the dynamics for both conditional covariance processes is little changed no matter whether one considers the ALSI or Mixed Portfolio in

---

<sup>97</sup>For the Mixed Portfolio, summing the squared daily portfolio returns is equivalent to calculating the monthly variance using the following expression:

$$variance_{MixedPortfolio} = w_{ALSI}^2 var(ALSI) + w_{ALBI}^2 var(ALBI) + 2w_{ALSI}w_{ALBI}cov(ALSI, ALBI)$$

Given that:

$$\begin{aligned} \sum r_{MixedPortfolio}^2 &= \sum (w_{ALSI} \times r_{ALSI} + w_{ALBI} \times r_{ALBI})^2 \\ &= w_{ALSI}^2 \sum r_{ALSI}^2 + w_{ALBI}^2 \sum r_{ALBI}^2 + 2w_{ALSI}w_{ALBI} \sum r_{ALSI}r_{ALBI} \\ &= w_{ALSI}^2 var(ALSI) + w_{ALBI}^2 var(ALBI) + 2w_{ALSI}w_{ALBI}cov(ALSI, ALBI) \end{aligned}$$

conjunction with the USDZAR forward<sup>98</sup>. The selection of the Scalar BEKK over a Full BEKK structure indicates that there are no volatility spillover effects from the ALSI or Mixed Portfolio to the USDZAR forward (or vice versa). The shape parameter is highly significant and greater than 2, indicating that the second moment does indeed exist.

Table 21 presents the DCC estimates. The finding of a statistically significant  $\hat{\gamma}$  for the ALSI and the Mixed Portfolio is unsurprising and in line with numerous studies that document an asymmetric volatility response in equity markets, (Nelson (1991), Engle and Ng (1993), Glosten et al. (1993)). The persistence of the USDZAR forward is significantly lower than the persistence of the ALSI and Mixed Portfolio. Defining the asymmetry ratio as  $A = (\alpha + \gamma)/\alpha$  we can see that the response of the ALSI and Mixed Portfolio conditional variance to returns lower than the mean is markedly higher, given an estimate for  $A$  of 7.92 and 5.10 respectively. Making direct comparisons between the estimates of  $A$  found in previous studies is complicated as there are differing time periods, indices, volatility models as well as return frequencies. In terms of daily returns, Nelson (1991) estimates  $A = 7.2$  for the CRSP value-weighted market index from 1962-1987. Bollerslev and Ole Mikkelsen (1996) find that  $A = 6$  for the S&P 500 index from 1961 to 1991 while Engle and Ng (1993) find  $A = 2.6$  for the Japanese TOPIX index during the period 1980-1988. In addition, Blair et al. (2001) estimate  $A = 8.5$  for the S&P 100 index from 1993-1998. Despite all the aforementioned obstacles one might make the tentative conclusion that the asymmetry ratio for the broader South African equity market is towards the upper end of previously reported empirical estimates.

The estimates for the GO-GARCH model in Table 22 have no natural interpretation, other than to say that the persistence of the unobserved factors is considerably less than the persistence of the observed data, vis-a-vis our DCC or even BEKK estimates. The RiskMetrics model doesn't require any parameter estimation or model selection, thus the only task remaining is to select an appropriate ARFIMA model for the following series: the log of realized USDZAR forward variance ( $\log(\text{realizedrand})$ ), the log of realized ALSI variance ( $\log(\text{realizedalsi})$ ) the log of realized Mixed Portfolio variance ( $\log(\text{realizedportfolio})$ ), the covariance between the USDZAR forward and ALSI ( $\text{covrandalsi}$ ) and the covariance between the USDZAR forward and the Mixed Portfolio ( $\text{covrandportfolio}$ ). All three log realized volatility series can be modelled by a parsimonious ARMA (1,1) model with very similar AR and MA estimates across series. These are contained in Table 23.

## 5.6 Multivariate forecasts

Having selected the best model from each MGARCH family, and an ARFIMA<sup>99</sup> model for the variances and covariances, we next turn to a comparison of competing forecasts from each of these models on the basis of statistical and economic measures. After calculating returns from monthly closing prices, we are left with a sample of 194 monthly observations, beginning in January 1997 and ending in February 2013. As in

<sup>98</sup>Note that both bivariate returns series are covariance stationary since  $(\hat{a}^2 + \hat{b}^2) < 1$

<sup>99</sup>We will continue to use the ARFIMA-RV nomenclature for our reduced-form modelling of the covariance matrix.

the univariate section, we calculate forecasts on the basis of a rolling sample of 132 (11 years) observations and then make a 1-step-ahead forecast. This process is repeated 62 times in order to obtain a series of out-of-sample forecasts spanning the period 1/2008 to 2/2013.

Figures 16 and 17 present the forecasts for the covariance matrix when the ALSI or Mixed Portfolio are paired with the USDZAR forward. There are three clear periods of heightened volatility in both of these figures. In October 2008, realized variances for the USDZAR, ALSI, Mixed Portfolio as well as the realized covariances between the USDZAR/ALSI and USDZAR/Mixed Portfolio, spike significantly and remain high for approximately 8-10 months depending on which sub-figure one is examining. In May 2010, there is another notable spike across all variances and covariances, but it is isolated and volatility drops in the following period. Finally, in August 2011 another multi-month period of increased volatility lasting around 5 months appears. In terms of magnitude, both the 2010 and 2011 periods are dwarfed by the extreme volatility in late 2008. Throughout each of these episodes all of the models were only able to react and not anticipate these sharp increases in volatility. In terms of the speed of adjustment to these swift volatility changes the DCC and ARFIMA-RV models are generally the quickest to react, the ARFIMA-RV model however seems to have an advantage when one considers the covariances. Overall the differences in the competing forecasts are mostly negligible during periods of lower volatility, it is only during the aforementioned high volatility periods that differences amongst forecasts become noticeable, as was the case in the univariate section. The DCC and BEKK model forecasts in particular track each other very closely while the GO-GARCH, RiskMetrics and ARFIMA-RV model seem to follow their own path. The RiskMetrics model has by far the smoothest series of forecasts, but the price for this is its slowness in reacting to changing market conditions, as can be seen in 2008-2009 where it under- and then subsequently overestimates volatility. A casual inspection of these figures hints at the ARFIMA-RV model's forecasting superiority, relative to the other models.

A more succinct way of comparing covariance matrix forecasts would be to utilize each element of the covariance matrix to calculate the betas or hedge ratios. Figures 18 presents the betas from each model as well as the "Realized Beta" which is simply calculated as  $beta = -\frac{realizedcovariance}{realizedrand}$ . The first insight gleaned from these figures would be the relative stability of the betas, both actual and forecasted, compared to the variances and covariances examined in Figures 16 and 17. Again, the beta from the RiskMetrics model is the least volatile by far. The realized beta for the USDZAR/ALSI mostly fluctuates around a mean of 0.4924 without any discernible trend as does the realized beta for the USDZAR/Portfolio around its trend of 0.3558. The ARFIMA-RV model, as was the case for the covariance matrix, appears to produce the most accurate forecasts, however, a more definitive answer is given in the next two subsections.

### 5.6.1 Statistical Evaluation of Multivariate forecasts

In this section we will use the SPA test of Hansen (2005) in conjunction with two robust multivariate loss functions found in Laurent et al. (2012) in order to test for superior forecasting performance amongst the BEKK, DCC, GO-GARCH, RiskMetrics and ARFIMA-RV models. The first loss function, the Euclidean

distance is defined by the weighting matrix  $\hat{\Lambda} = \mathbf{I}_3$  and treats forecasts of the variance and covariance equally. It could be argued that the covariance carries greater weight in multivariate applications and for this reason we also include the Frobenius distance, defined by the weighting matrix  $\hat{\Lambda} = \text{diag}([1 \ 1 \ 2])$  that punishes over or under-prediction of the covariance more harshly.

As in the univariate section, each individual model will be the benchmark model on one occasion, and the remaining four models will form a  $62 \times 4$  matrix whose columns hold the Euclidean or Frobenius distance for each of the alternative models.

Table 24 presents the SPA test results for the ALSI and Mixed Portfolio, based on the Euclidean distance. For the ALSI, we cannot reject the null that no alternative model outperforms the benchmark, when that benchmark is either the BEKK, DCC, GO-GARCH or ARFIMA-RV model. If the RiskMetrics model is the benchmark, then we can reject the null of no outperformance at the 5% significance level. A p-value of 1 for the DCC model suggests that it may be the best model, however on the whole the results indicate that with exception of the RiskMetrics, the performances of all the models are similar, sufficiently so that we cannot reject the SPA test's null hypothesis.

The picture is quite different when one considers the Mixed Portfolio. In this case we can reject the null hypothesis at the 5% level for the BEKK and RiskMetrics model or at the 10% level for the DCC and GO-GARCH model. The ARFIMA-RV model however has a p-value of 1, providing the strongest indication of statistical outperformance yet. The background events and dramatic increases in market volatility in late 2008 were outliers in the truest sense of the word, and removing this period from the sample provides additional insight into the performance of each these models during different states of the market. Table 25 provides the p-values of the SPA test but with the three most volatile months in the sample removed, those being September, October and November 2008. After removing the outliers, the ARFIMA-RV model is the clear outperformer, based on both the ALSI and Mixed Portfolio, with p-values of 1 in both cases. As was the case for the full sample, the p-values suggest a larger disparity in performance for the Mixed Portfolio, as opposed to the ALSI, given that we are unable to reject the null of outperformance for the BEKK and DCC model at the ten percent level for the ALSI portfolio.

Moving on to the Frobenius multivariate loss function, which is given in Tables 26 and 27. The general trend is very much the same; the ARFIMA-RV model is clearly superior in the Mixed Portfolio case but its performance cannot be distinguished from a number of other models in the ALSI scenario. With the exception of the RiskMetrics model, one is unable to reject the null of outperformance for each of the remaining models. Over the complete sample the DCC model seems to be the strongest performer for the ALSI once again. Removing the three volatile months in 2008 achieves the exact same result as before, in that the ARFIMA-RV model usurps the DCC model when one considers the ALSI, although we cannot reject the null that no model outperforms the BEKK or DCC model at the 5% level.

Some preliminary conclusions that we can draw from these tests. Firstly, in terms of a pure equity portfolio, there is strong evidence for the superiority of the realized volatility approach over traditional

MGARCH models, in ‘normal’ market conditions. Under extreme market conditions, as in 2008, the DCC model may appear to outperform the ARFIMA-RV model, so much so that its p-value decreased from 1 to 0.3634 upon removing those outliers from the sample, as measured by the Euclidean distance. In the case of the Frobenius distance, the DCC model’s p-value decreases from 1 to 0.1002. When one considers the Mixed Portfolio however, the ARFIMA-RV approach is the best irrespective of the sample or loss function used.

### 5.6.2 Economic Evaluation of Multivariate forecasts

Having compared the MGARCH and ARFIMA-RV models on a pure statistical basis, we now make the same comparison using the hedging framework outlined in Section Four. Our hedger uses each element of the covariance forecast from the set of models in order to determine the quantity of USDZAR forwards he should buy in order to reduce the variance of his portfolio, which has predetermined weights ( $w_{ALSI}, w_{ALBI}$ ) that are fixed at (1,0) or (0.6, 0.4) throughout the entire sample.

Table 28 gives the average ex-post utility that an investor with only equity holdings would have received over the 62 months had he left his predetermined portfolio, either unhedged or hedged, based on the conditional hedge ratios obtained from either the MGARCH models, the ARFIMA-RV model or a simple OLS regression of his portfolio returns on the USDZAR forward<sup>100</sup>. Given that the investor’s utility is a function of returns one must interpret this table with caution as we have not made any attempts to model or predict returns. Therefore the rankings based on utility are largely sample-dependent, particularly for lower levels of risk aversion. However, for higher levels of risk aversion, more specifically for  $\gamma \geq 4$  the ARFIMA-RV model is the highest ranked model<sup>101</sup> and this is largely due to the significant variance reduction that the investor is able to achieve using this model. Table 29 gives the annualized variance of the investor’s portfolio, providing further confirmation of this. Conditional hedging using an ARFIMA-RV derived hedge ratio would have reduced the variance of the investor’s holdings by more than 32%. In fact, any form of hedging, whether it is using the conventional hedge or one of the conditional hedges reduces the variance. Note however that the gains from conditional hedging are on average far greater than the conventional OLS hedge. Within the conditional hedging group the GO-GARCH model is a notable laggard, achieving a marginally higher risk reduction than the OLS hedge.

Tables 30 and 31 present the ex-post utility and annualized variance for the Mixed Portfolio. Without placing too much emphasis on utility, the ARFIMA-RV model is again the highest ranked model for low to moderate levels of risk aversion, when  $\gamma = 6$  however, the marginally greater risk reduction achieved by the RiskMetrics model begins to have an impact on the rankings<sup>102</sup>. In Table 31 we see that the risk reduction achieved for the Mixed Portfolio is far more uniform across both the conditional and conventional hedges. The lowest risk reduction is provided by the OLS hedge, yet it only lags behind the worst performing

<sup>100</sup>The latter’s inclusion will facilitate a comparison between the conventional means of hedging and conditional hedging.

<sup>101</sup>An assumption of  $\gamma \geq 4$  is largely supported by empirical studies. Chou (1988) estimate  $\gamma = 4.5$ , Poterba and Summers (1986) find  $\gamma = 3.5$  while Grossman and Shiller (1981) find that  $\gamma = 6$ .

<sup>102</sup>Although the BEKK model provided a larger risk reduction its returns were lower than those of the RiskMetrics model.

MGARCH model (DCC) by approximately 0.4%. Amongst the conditional hedges the BEKK, RiskMetrics and ARFIMA-RV model are able to achieve a sizable risk reduction of 40%, which is somewhat surprising given the already reduced risk of the Mixed Portfolio, following the inclusion of bonds.

Thus far no mention has been made of transaction costs, and a definitive answer to the question of whether the investor should hedge or not can only be given after accounting for the spread that the investor would have to pay when opening and closing his/her futures position. Transaction costs are usually measured using the relative spread, which accounts for exchange rates of varying magnitudes and is calculated as  $100 * \ln(ASK/BID)$ . Using the median relative spread for the South African Rand reported in Hassan and Smith (2011), of 0.8029%, we can make a very rough adjustment for transaction costs by reducing the return to the hedged position by  $-\beta \times 0.8029$ . Thus, conditional hedges that prescribe larger USDZAR forward positions will incur greater transaction costs, thereby reducing their returns by a disproportionate amount. Given that the goal of the hedging exercise was solely risk minimization and that the monthly variance is unchanged when including transaction costs we have not reported the rankings of the respective models inclusive of those costs. The DCC model is the best performer after making this ‘back-of-the-envelope’ adjustment for transaction costs<sup>103</sup>. An investor would be willing to pay up to 23 basis points monthly for his/her ALSI portfolio to be hedged based on the conditional covariance matrix from the DCC model, or 9 basis points monthly for his Mixed Portfolio to be hedged using the same model<sup>104</sup>. The investor is willing to pay such a price for this hedging due to the large reduction in the variance of his holdings, while leaving his returns relatively unchanged, even after accounting for transaction costs.

To summarize; for empirically plausible levels of risk aversion the ARFIMA-RV approach is the highest ranked model for both the ALSI and Mixed Portfolio, although this ranking is affected by the sample-dependent returns. In terms of pure risk reduction a conditional hedging strategy using traditional MGARCH models is able to provide comparable performance, even exceeding the risk reduction of the ARFIMA-RV model for the Mixed Portfolio. For the investor, irrespective of the composition of his portfolio or risk tolerance, the gains from hedging are significant.

## 6 Conclusion and directions for future research.

In this paper we examined the properties of daily realized USDZAR spot rate volatility that was obtained using the MA(1)-based realized variance estimator of Hansen et al. (2008). We found that the results accorded very well with the findings of Andersen et al. (2001), where the realized variance was approximately log-normally distributed, but with a higher probability of extremely volatile days. In addition, the realized

<sup>103</sup>Unsurprisingly the DCC model, with the lowest average hedge ratio, incurred the lowest transaction costs while still providing a significant amount of risk reduction.

<sup>104</sup>Given that the investor’s average monthly utility is given by  $r - \gamma \times var(r)$ , one can determine the maximum amount that the investor would be willing to pay (in basis points) for the hedge, by decreasing his returns from the hedged portfolio until they yield the same amount of utility as his unhedged portfolio.

USDZAR variance was found to be a fractionally integrated process. We then examined the properties of the covariance matrix for a bivariate returns series consisting of the 1-month USDZAR forward and either monthly ALSI returns or the monthly returns from a 60/40 equity and bond portfolio. The realized covariance matrix was calculated using the traditional realized covariance estimator of Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002). It was found that the realized variances were again approximately log-normal but no longer possessed the long memory property. In contrast, the realized covariances were found to be fractionally integrated, significantly left-skewed, and leptokurtic.

Based on these insights, we then fitted reduced-form time series models to the realized variance and realized covariance and compared their 1-step-ahead forecasting performance to a number of univariate and multivariate GARCH models respectively. Using "proxy-robust" statistical loss functions allied with a test for Superior Predictive Ability, the reduced-form models were generally found to be superior to the GARCH competitors. The traditional GARCH models were not able to explain variation in the realized variance over and above that which was explained by the ARFIMA-RV model. Thus, the assertion that the direct modelling of realized volatility is superior to the traditional GARCH literature is largely supported here. All is not lost for the GARCH models however, as the HEAVY model (which is a blend of the GARCH and realized volatility literature) was able to provide forecasting performance comparable to that of the ARFIMA-RV model. On the whole, our univariate analysis would lead us to conclude that the significant advantage of the reduced-form time series model is due to the precision of the volatility proxy.

The finding of a negative asymmetric volatility response for daily USDZAR volatility could potentially be explained by the so called 'Risk-on Risk-off' phenomenon. Campbell et al. (2007) find an inverse relationship between world equity prices and the Dollar, Euro and Swiss franc. They posit that this represents a "flight to quality" by investors who regard those currencies and assets denominated in those currencies as safe havens. However, these funds do not appear out of thin air, but instead are being transferred from one country to another. Given South Africa's status as an emerging market economy and the perceived risk associated with investing in South African bonds or equities, portfolio flows from South African rand assets and into safe-haven currencies are likely to increase when there is greater global risk-aversion and more volatility. Hence, this could possibly explain why volatility is higher following a depreciation of the currency versus an appreciation. Another possibility is the unwinding of carry trade positions where the rand is being used as the target currency<sup>105</sup>. The profitability of these trades relies on the failure of Uncovered Interest Rate parity to hold. The result is that there is a feedback effect whereby a depreciation of the rand makes the carry trade more risky and leads to an unwinding of these positions, further exacerbating a rand depreciation and manifesting greater volatility.

In terms of the multivariate application the conditional covariance matrices were also evaluated on an economic basis whereby an investor would hedge the currency risk of his portfolio using the conditional covariance matrix. This exercise was predicated on a contemporaneous relationship in the mean of USDZAR

<sup>105</sup>See Hassan and Smith (2011), Brunnermeier et al. (2008).



forwards and the mean of the ALSI or Mixed Portfolio. Interestingly, there was no relationship between their second moments, as evidenced by the selection of the Scalar BEKK over the Full BEKK. In other words there were no volatility spillover effects between the USDZAR forward rate and the ALSI or Mixed Portfolio.

There was evidence to suggest that the volatility in the South African equity market responded differently to negative or positive returns. Following returns lower than expected in period  $t$ , the ALSI would experience higher volatility in period  $t+1$  than if the previous period's returns had been positive. While this is a common finding in equity markets, the magnitude of this asymmetry was quite high compared to previous findings.

On the whole the question of whether the direct modelling approach is superior for multivariate applications is less clear cut. The need to compare models using a variety of loss functions, both statistical and economic was very apparent. As an illustration, the RiskMetrics model was unquestionably the worst performer in the statistical evaluation, however, it was able to provide a very competitive amount of risk reduction within our economic evaluation. The results from the hedging example do not preclude the possibility of the ARFIMA model being outperformed by an MGARCH model for a specific loss function. Although not central to this paper, the implications of the hedging example are that there are tangible economic benefits for an investor who wants to reduce the variance of his portfolio using a hedge calculated from the conditional covariance matrix. The gains from hedging were significantly reduced portfolio variance, without sacrificing much in the way of returns, even after accounting for transaction costs.

In both the statistical and economic evaluation the DCC model was generally the best performer out of the MGARCH models, highlighting the importance of correlations in multivariate applications and the benefit of parameterizing them directly. Of course the inclusion of an asymmetric volatility response parameter for the investor's portfolio could have been the source of the DCC model's edge. However, the ARFIMA-RV model could easily incorporate this asymmetry by including a dummy variable equal to one when the prior residual is negative and zero otherwise.

In terms of other topics for future research, there are a number of avenues that could be further explored. The approach taken here was to focus solely on the 1-step ahead forecast but the analysis could readily be extended to include the traditional 5, 10 and 22-step forecasts. Additionally the MA(1) based realized variance is but one of a number of microstructure noise robust volatility estimators and one could assess the forecasting performance of ARFIMA-RV models using alternative volatility estimators such as the realized kernel of Barndorff-Nielsen et al. (2008), the TTSE of Zhang et al. (2005) or even the range-based estimator of Alizadeh et al. (2002). Other bilateral exchange rates between South Africa and its trading partners such as the Rand price of the Euro (EURZAR), British Pound (GBPZAR) or Japanese Yen (JPYZAR) could be included in the future in order to compare the performance of the models in a higher-dimensional setting. If the hedging exercise were repeated with these currencies included then one would be able to ascertain if hedging the exposure from multiple currencies brings about further risk reduction or whether focusing only on the USDZAR is sufficient.

## References

- Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* 18(2), 351–416.
- Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics* 160(1), 160–175.
- Alexander, C. and A. Chibumba (1997). Orthogonal garch: an empirical validation in equities, foreign exchange and interest rates. *School of Mathematical Sciences Discussion Paper, Sussex University*.
- Alizadeh, S., M. W. Brandt, and F. X. Diebold (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance* 57(3), 1047–1091.
- Andersen, T. and T. Bollerslev (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of empirical finance* 4(2), 115–158.
- Andersen, T. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 885–905.
- Andersen, T., T. Bollerslev, P. Christoffersen, and F. Diebold (2006). Volatility and correlation forecasting. *Handbook of economic forecasting* 1, 777–878.
- Andersen, T., T. Bollerslev, and F. Diebold (2002). Parametric and nonparametric volatility measurement.
- Andersen, T., T. Bollerslev, and F. Diebold (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics* 89(4), 701–720.
- Andersen, T., T. Bollerslev, F. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics* 61(1), 43–76.
- Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (1999). Realized volatility and correlation.
- Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica* 71(2), 579–625.
- Andersen, T., T. Bollerslev, and N. Meddahi (2005). Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica* 73(1), 279–296.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American statistical association* 96(453), 42–55.
- Anderson, R. W. and J.-P. Danthine (1981). Cross hedging. *The Journal of Political Economy*, 1182–1196.
- Bai, X., J. Russell, and G. Tiao (2000). *Beyond Merton's utopia: effects of non-normality and dependence on the precision of variance estimators using high-frequency financial data*. Ph. D. thesis, University of Chicago, Graduate School of Business.

- Baillie, R. and T. Bollerslev (1991). Intra-day and inter-market volatility in foreign exchange rates. *The Review of Economic Studies* 58(3), 565–585.
- Baillie, R. and T. Bollerslev (2002). The message in daily exchange rates. *Journal of Business & Economic Statistics* 20(1), 60–68.
- Baillie, R., T. Bollerslev, and H. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74(1), 3–30.
- Baillie, R. and R. Myers (1991). Bivariate garch estimation of the optimal commodity futures hedge. *Journal of Applied Econometrics* 6(2), 109–124.
- Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of econometrics* 73(1), 5–59.
- Bandi, F. and J. Russell (2008). Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies* 75(2), 339–369.
- Bandi, F. M. and J. R. Russell (2006). Comment. *Journal of Business & Economic Statistics* 24(2), 167–173.
- Bandi, F. M. and J. R. Russell (2011). Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. *Journal of Econometrics* 160(1), 145–159.
- Barndorff-Nielsen, O. and N. Shephard (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* 17(5), 457–477.
- Barndorff-Nielsen, O. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2(1), 1–37.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2004). Regular and modified kernel-based estimators of integrated variance: The case with independent noise. *Economics Group, Nuffield College, University of Oxford, Economics Papers*.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* 76(6), 1481–1536.
- Bauwens, L., S. Laurent, and J. Rombouts (2006). Multivariate garch models: a survey. *Journal of Applied Econometrics* 21(1), 79–109.
- Berndt, E., B. Hall, and R. Hall (1974). Estimation and inference in nonlinear structural models. In *Annals of Economic and Social Measurement, Volume 3, number 4*, pp. 103–116. NBER.
- Black, F. (1976). Studies of stock price volatility changes.
- Black, F. (1990). Equilibrium exchange rate hedging. *Journal of Finance*, 899–907.
- Blair, B. J., S.-H. Poon, and S. J. Taylor (2001). Modelling s&p 100 volatility: The information content of stock returns. *Journal of Banking & Finance* 25(9), 1665–1679.

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3), 307–327.
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 542–547.
- Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The Review of Economics and Statistics*, 498–505.
- Bollerslev, T., R. Chou, and K. Kroner (1992). Arch modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* 52(1), 5–59.
- Bollerslev, T., R. Engle, and D. Nelson (1994). Arch models.
- Bollerslev, T., R. Engle, and J. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *The Journal of Political Economy*, 116–131.
- Bollerslev, T. and H. Ole Mikkelsen (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* 73(1), 151–184.
- Bollerslev, T. and J. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews* 11(2), 143–172.
- Boswijk, H. P. and R. Van der Weide (2006). Wake me up before you go-garch.
- Box, G. and G. Jenkins (1970). *Time Series Analysis: Forecasting and Control*. Holden-D. iv, San Francisco.
- Bregman, L. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics* 7(3), 200–217.
- Brunnermeier, M. K., S. Nagel, and L. H. Pedersen (2008). Carry trades and currency crashes. Technical report, National Bureau of Economic Research.
- Burnside, C., M. Eichenbaum, I. Kleshchelski, and S. Rebelo (2006). The returns to currency speculation. Technical report, National Bureau of Economic Research.
- Campbell, J. Y., K. S.-d. Medeiros, and L. M. Viceira (2007). Global currency hedging. Technical report, National Bureau of Economic Research.
- Caporin, M. and M. McAleer (2009). Do we really need both bekk and dcc? a tale of two covariance models. *A Tale of Two Covariance Models (February 5, 2009)*.
- Chen, N.-F., R. Roll, and S. A. Ross (1986). Economic forces and the stock market. *Journal of business*, 383–403.
- Chou, R. Y. (1988). Volatility persistence and stock valuations: Some empirical evidence using garch. *Journal of Applied Econometrics* 3(4), 279–294.

- Christie, A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial Economics* 10(4), 407–432.
- Corsi, F., G. Zumbach, U. Muller, and M. Dacorogna (2003). Consistent high-precision volatility from high-frequency data. *Economic Notes* 30(2), 183–204.
- Dacorogna, M. and R. M. Gencay (2001). *An introduction to high frequency finance*. Academic Press, London.
- Dacorogna, M., U. Müller, R. Nagler, R. Olsen, and O. Pictet (1993). A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance* 12(4), 413–438.
- Diebold, F. X. and R. S. Mariano (1994). Comparing predictive accuracy. *Journal of Business & economic statistics* 20(1).
- Ding, Z. and R. F. Engle (2001). Large scale conditional covariance matrix modeling, estimation and testing. *ACADEMIA ECONOMIC PAPERS* 29(2), 157–184.
- Ding, Z., C. Granger, and R. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1(1), 83–106.
- Doornik, J. A. (2009). An object-oriented matrix programming language ox 6.
- Duffee, G. (1995). Stock returns and volatility a firm-level analysis. *Journal of Financial Economics* 37(3), 399–420.
- Engel, C. (1996). The forward discount anomaly and the risk premium: A survey of recent evidence. *Journal of Empirical Finance* 3(2), 123–192.
- Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.
- Engle, R. (2001). Garch 101: An introduction to the use of arch/garch models in applied econometrics.
- Engle, R. (2002a). Dynamic conditional correlation. *Journal of Business & Economic Statistics* 20(3), 339–350.
- Engle, R. (2002b). New frontiers for arch models. *Journal of Applied Econometrics* 17(5), 425–446.
- Engle, R. and T. Bollerslev (1986). Modelling the persistence of conditional variances. *Econometric reviews* 5(1), 1–50.
- Engle, R. and R. Colacito (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics* 24(2), 238–253.
- Engle, R. and G. Gonzalez-Rivera (1991). Semiparametric arch models. *Journal of Business & Economic Statistics* 9(4), 345–359.

- Engle, R., D. Hendry, and D. Trumble (1985). Small-sample properties of arch estimators and tests. *Canadian Journal of Economics*, 66–93.
- Engle, R. and K. Kroner (1995). Multivariate simultaneous generalized arch. *Econometric theory* 11(01), 122–150.
- Engle, R., D. Lilien, and R. Robins (1987). Estimating time varying risk premia in the term structure: the arch-m model. *Econometrica: Journal of the Econometric Society*, 391–407.
- Engle, R. and J. Mezrich (1996). Garch for groups: A round-up of recent developments in garch techniques for estimating correlation. *RISK-LONDON-RISK MAGAZINE LIMITED*- 9, 36–40.
- Engle, R., V. Ng, and M. Rothschild (1990). Asset pricing with a factor-arch covariance structure: empirical estimates for treasury bills. *Journal of Econometrics* 45(1), 213–237.
- Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* 131(1), 3–27.
- Engle, R. F. and V. K. Ng (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance* 48(5), 1749–1778.
- Fama, E. (1965). The behavior of stock-market prices. *Journal of Business*, 34–105.
- Fleming, J., C. Kirby, and B. Ostdiek (2003). The economic value of volatility timing using realized volatility. *Journal of Financial Economics* 67(3), 473–509.
- Foster, D. P. and D. B. Nelson (1994). Continuous record asymptotics for rolling sample variance estimators.
- French, K., G. Schwert, and R. Stambaugh (1987). Expected stock returns and volatility. *Journal of Financial Economics* 19(1), 3–29.
- Galati, G., A. Heath, and P. McGuire (2007). Evidence of carry trade activity. *BIS Quarterly Review* 3, 27–41.
- Garman, M. and M. Klass (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, 67–78.
- Geweke, J. and S. Porter-Hudak (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4(4), 221–238.
- Glen, J. and P. Jorion (1993). Currency hedging for international portfolios. *The Journal of Finance* 48(5), 1865–1886.
- Glosten, L., R. Jagannathan, and D. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* 48(5), 1779–1801.
- Goodhart, C. and L. Figliuoli (1991). Every minute counts in financial markets. *Journal of International Money and Finance* 10(1), 23–52.

- Granger, C. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1(1), 15–29.
- Grossman, S. J. and R. J. Shiller (1981). The determinants of the variability of stock market prices.
- Guillaume, D., M. Dacorogna, R. Davé, U. Müller, R. Olsen, and O. Pictet (1997). From the bird’s eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and stochastics* 1(2), 95–129.
- Hansen, L. P. and R. J. Hodrick (1980). Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *The Journal of Political Economy*, 829–853.
- Hansen, P. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics* 23(4), 365–380.
- Hansen, P. and A. Lunde (2004). An unbiased measure of realized variance. *Available at SSRN 524602*.
- Hansen, P. and A. Lunde (2005). A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of Applied Econometrics* 20(7), 873–889.
- Hansen, P. and A. Lunde (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics* 24(2), 127–161.
- Hansen, P. R., Z. Huang, and H. H. Shek (2012). Realized garch: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*.
- Hansen, P. R., J. Large, and A. Lunde (2008). Moving average-based estimators of integrated variance. *Econometric Reviews* 27(1-3), 79–111.
- Hassan, S. and S. Smith (2011). The rand as a carry trade target: Risk, returns and policy implications. Working Papers 235, Economic Research Southern Africa.
- He, C. and T. Teräsvirta (2004). An extended constant conditional correlation garch model and its fourth-moment structure. *Econometric Theory* 20(05), 904–926.
- Hsieh, D. (1989). Modeling heteroscedasticity in daily foreign-exchange rates. *Journal of Business & Economic Statistics* 7(3), 307–317.
- Hurst, H. (1950). Long-term storage capacity of reservoirs. *American Society of Civil Engineering* 76.
- Koopman, S. J., B. Jungbacker, and E. Hol (2005). Forecasting daily variability of the s&p 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* 12(3), 445–475.
- Kroner, K. and J. Sultan (1993). Time-varying distributions and dynamic hedging with foreign currency futures. *Journal of Financial and Quantitative Analysis* 28(4).
- Laurent, S., J. Rombouts, and F. Violante (2012). On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics*.

- Ledoit, O., P. Santa-Clara, and M. Wolf (2003). Flexible multivariate garch modeling with an application to international stock markets. *Review of Economics and Statistics* 85(3), 735–747.
- Lee, G. and R. Engle (1993). A permanent and transitory component model of stock return volatility. Available at SSRN 5848.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 394–419.
- Marcucci, J. (2006). Forecasting stock market volatility with regime-switching garch models. *Studies in Nonlinear Dynamics & Econometrics* 9(4).
- McAleer, M. and M. C. Medeiros (2008). Realized volatility: A review. *Econometric Reviews* 27(1-3), 10–45.
- Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8(4), 323–361.
- Metrics, R. (1997). Technical document. *JP Morgan*.
- Mincer, J. and V. Zarnowitz (1969). The evaluation of economic forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, pp. 1–46. NBER.
- Muller, U., M. Dacorogna, R. Olsen, O. Pictet, M. Schwarz, and C. Morgenegg (1990). Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking & Finance* 14(6), 1189–1208.
- Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.
- Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society*, 703–708.
- Noureldin, D., N. Shephard, and K. Sheppard (2011). Multivariate high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics*.
- Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, 61–65.
- Patton, A. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160(1), 246–256.
- Patton, A., D. N. Politis, and H. White (2009). Correction to automatic block-length selection for the dependent bootstrap by d. politis and h. white. *Econometric Reviews* 28(4), 372–375.
- Patton, A. and K. Sheppard (2009). Evaluating volatility and correlation forecasts. *Handbook of Financial Time Series*, 801–838.
- Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association* 89(428), 1303–1313.



- Politis, D. N. and H. White (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* 23(1), 53–70.
- Poterba, J. M. and L. H. Summers (1986). The persistence of volatility and stock market fluctuations. *The American Economic Review*.
- Ross, S. (1973). *The arbitrage theory of capital asset pricing*. Rodney L. White Center for Financial Research, University of Pennsylvania, The Wharton School.
- Shephard, N. and K. Sheppard (2010). Realising the future: forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics* 25(2), 197–231.
- Silvennoinen, A. and T. Teräsvirta (2009). Multivariate garch models. *Handbook of Financial Time Series*, 201–229.
- Solnik, B. H. (1974). An equilibrium model of the international capital market. *Journal of Economic Theory* 8(4), 500–524.
- Van der Weide, R. (2002). Go-garch: a multivariate generalized orthogonal garch model. *Journal of Applied Econometrics* 17(5), 549–564.
- West, K. D., H. J. Edison, and D. Cho (1993). A utility-based comparison of some models of exchange rate volatility. *Journal of International Economics* 35(1), 23–45.
- White, H. (2003). A reality check for data snooping. *Econometrica* 68(5), 1097–1126.
- Yang, D. and Q. Zhang (2000). Drift-independent volatility estimation based on high, low, open, and close prices. *The Journal of Business* 73(3), 477–492.
- Zakoian, J. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control* 18(5), 931–955.
- Zhang, L., P. A. Mykland, and Y. Ait-Sahalia (2005). A tale of two time scales. *Journal of the American Statistical Association* 100(472), 1394–1411.
- Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics* 14(1), 45–52.

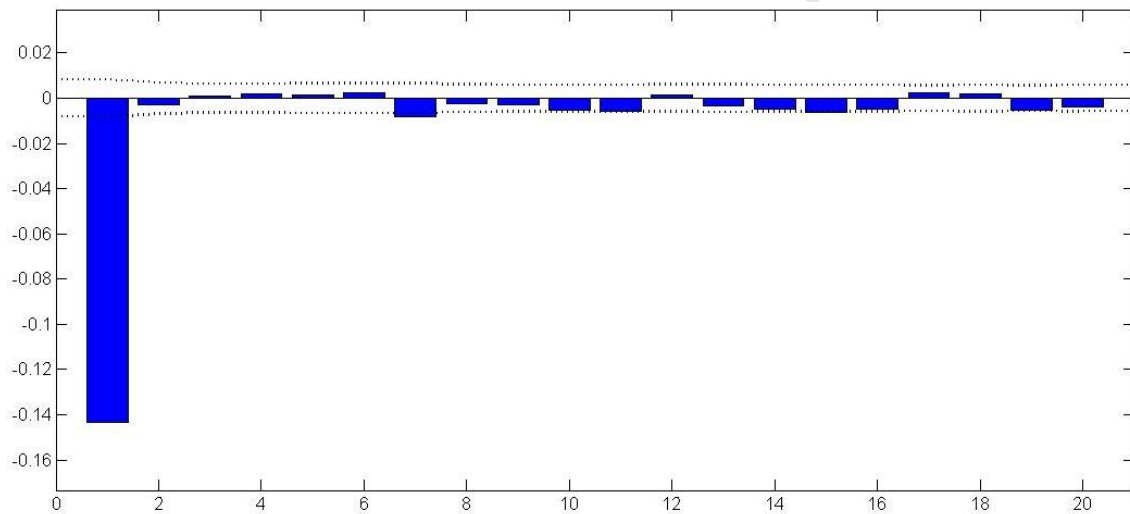
# Tables and Figures:

**Table 1: Descriptive Statistics for 5 minute USDZAR returns**

Mean	Standard Deviation	Skewness	Kurtosis	Min	Max	JB	AC(1)	LM(20)	Q <sup>2</sup> (20)
5.689e-07	9.302e-04	-0.2978	30.7032	-0.0394	0.0236	1.17e+07*	-0.1431*	24812*	60861*

Notes: The 5 minute returns were calculated as the difference in the log closing prices. The sample period begins the 1/01/2008 and ends the 1/01/2013 for a total of 368543 observations. The Jarque-Bera is the normality test which is  $\chi^2(2)$  distributed under the null of a normal distribution, here we reject the null of normality at the 5% level (indicated by \*). AC(1) is an estimate of the sample first-order autocorrelation which is significant using robust standard errors. LM(20) is the Lagrange Multiplier test statistic for ARCH effects up to the 20th lag which under the null hypothesis of no ARCH effects is  $\chi^2(20)$  distributed. Q<sup>2</sup>(20) is the Ljung-Box Q-statistic that tests for serial autocorrelation in the squared returns up to the 20th lag. Under the null of no serial autocorrelation this test is also  $\chi^2(20)$  distributed, thus we can safely reject the null for both of these tests at the 5% significance level since both of them exceed the critical value of 31.41.

**Figure 1: Sample ACF for 5 minute USDZAR returns**



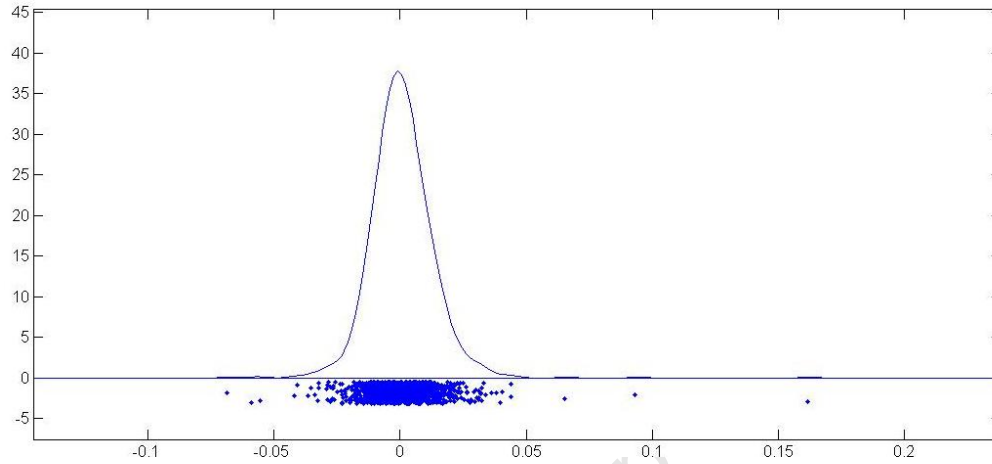
Notes: The figure shows the sample ACF of 5 minute USDZAR returns. The dashed lines represent 95% confidence intervals constructed using robust standard errors.

**Table 2: Descriptive Statistics for daily USDZAR returns**

Mean	Standard Deviation	Skewness	Kurtosis	Min	Max	$JB$	$AC(1)$	$LM(20)$	$Q^2(20)$
1.842e-04	0.0128	1.8394	25.8625	-0.0687	0.1617	2.84e+04*	-0.0172	207.87*	308.78*

*Notes: See Table 1 notes.*

**Figure 2: Density plot for daily USDZAR returns**



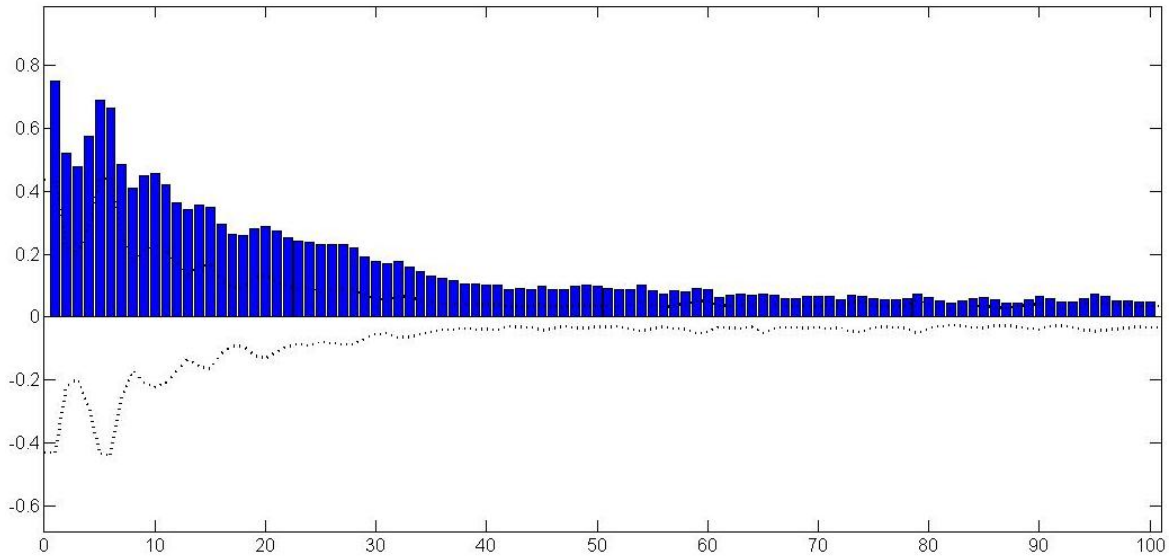
*Notes: The figure shows the unconditional density for daily USDZAR returns.*

**Table 3: Descriptive Statistics for realized variance**

Mean	Standard Deviation	Skewness	Kurtosis	Min	Max	<i>JB</i>	<i>AC</i> (1)	<i>Q</i> (20)
1.7434e-04	2.9582e-04	9.3558	123.4507	2.0983e-05	0.0052	7.87e+05*	0.7502*	5383*

*Notes: See Table 1 notes. The *Q*(20) is the Ljung-Box *Q*-stat calculated for the *MA*(1) realized variance.*

**Figure 3: Sample ACF-MA(1) realized variance (daily)**

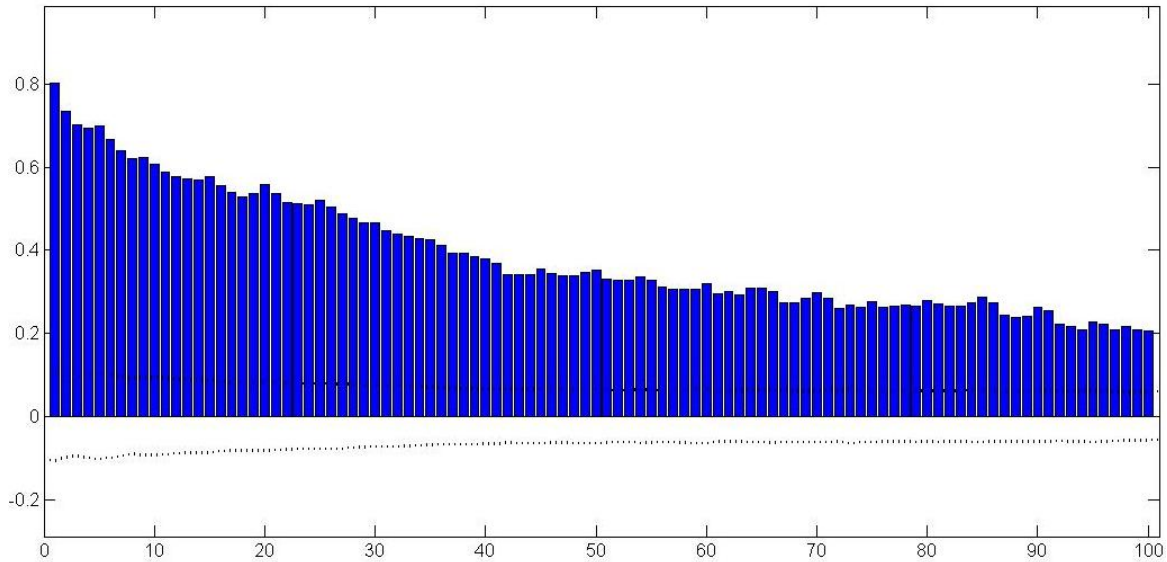


*Notes: The figure shows the sample ACF of daily realized volatility. The dashed lines represent 95% confidence intervals constructed using robust standard errors.*

**Table 4: Descriptive Statistics for log(RV)**

Mean	Standard Deviation	Skewness	Kurtosis	Min	Max	<i>JB</i>	<i>AC</i> (1)	<i>Q</i> (20)
-9.0534	0.7738	0.8785	4.6233	-10.7718	-5.2627	302.5768*	0.8039*	10057*

Notes: See Table 3 notes.

**Figure 4: Sample ACF-log(RV)**

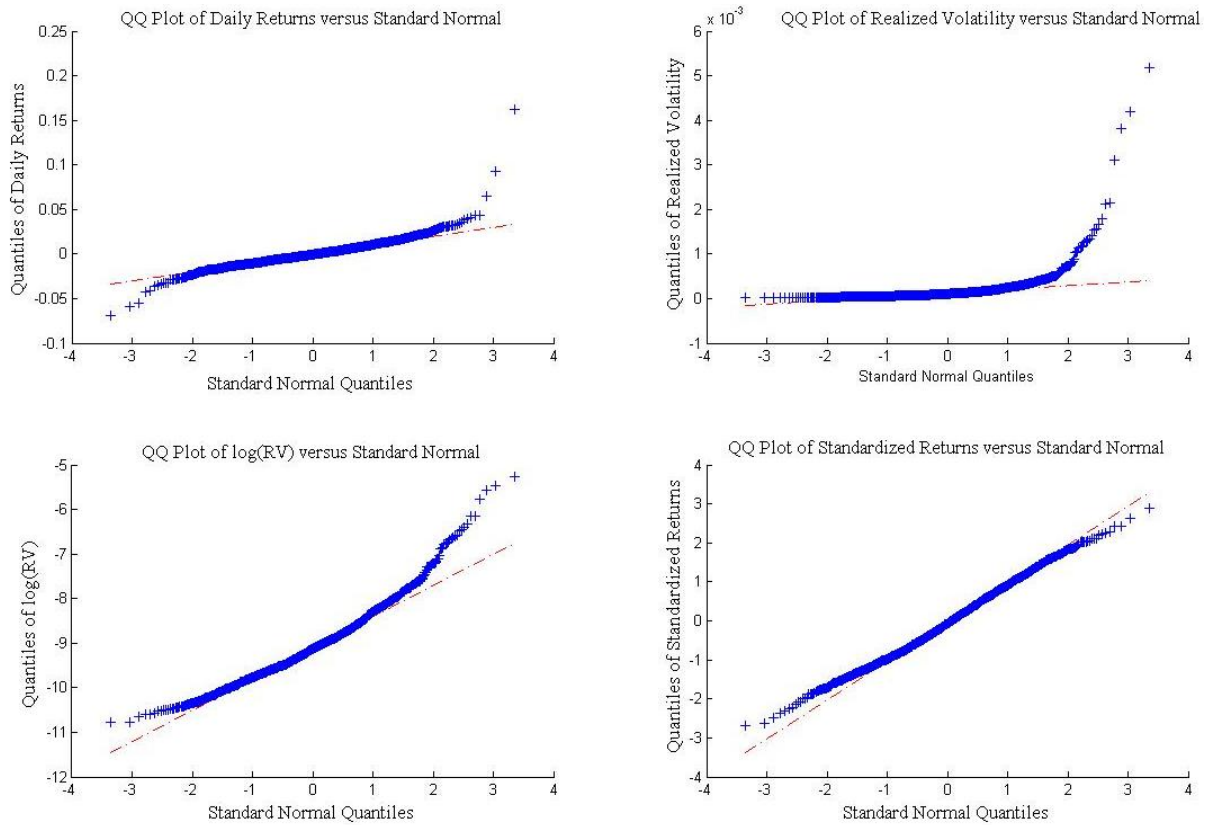
Notes: The figure shows the sample ACF of log(RV). The dashed lines represent 95% confidence intervals constructed using robust standard errors.

**Table 5: Descriptive Statistics for Standardized Daily Returns**

Mean	Standard Deviation	Skewness	Kurtosis	Min	Max	$JB$	$AC(1)$	$Q(20)$
-0.0334	0.9135	0.1523	2.6329	-3.6782	2.8956	12.249*	0.0378	16.807

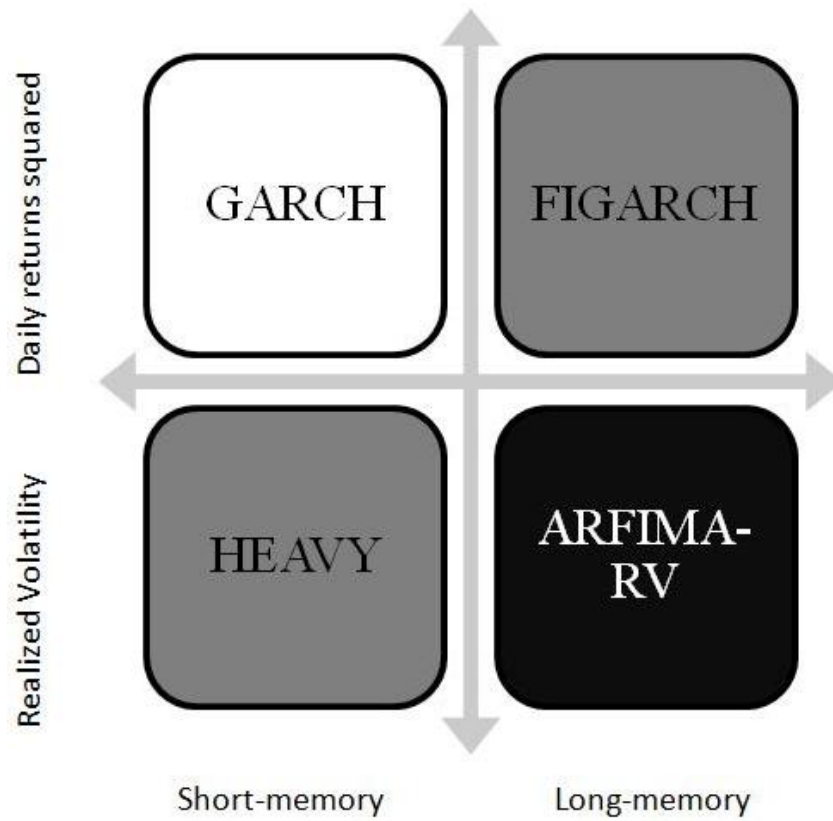
*Note: See Table 3 notes.*

**Figure 5: Quantile-Quantile plots for daily returns, realized variance,  $\log(RV)$  and standardized returns**



*Notes: These figures plot the quantiles of a Standard Normal distribution against the quantiles of daily returns, MA(1) realized variance,  $\log(RV)$  and standardized daily returns.*

**Figure 6: Model Schema**



*Note: The figure illustrates how the univariate models relate to one another in terms of the volatility proxy used (vertical axis), and the amount of memory in the volatility process (horizontal axis)*

**Table 6: Parameter estimates for the GJR-GARCH (1,1) model**

Parameter	Distribution		
	Normal	Student's-t	GED
$\hat{\omega}$	0.026279e-4 (0.0096291e-4)	0.027196e-4 (0.011244e-4)	0.027632e-4 (0.010112e-4)
$\hat{\alpha}$	0.142104 (0.042554)	0.101017 (0.030071)	0.126689 (0.034825)
$\hat{\beta}$	0.905668 (0.018534)	0.916978 (0.022843)	0.909166 (0.019853)
$\hat{\gamma}$	-0.139013 (0.049909)	-0.088820 (0.030908)	-0.120658 (0.039281)
$\hat{\nu}$		11.318375 (4.1056)	1.597541 (0.15753)
$\hat{\phi}$	0.978266	0.973584	0.975526
SBIC	-7913.12	-7950.18	-7928.07

*Note: The table presents the MLE for a GJR-GARCH(1,1,1) model for the conditional variance with three possible error distributions. The conditional variance for each model is given by  $h_t = \omega + \alpha(e_{t-1}^2) + I_{e_{t-1} < 0} \gamma(e_{t-1}^2) + \beta(h_{t-1})$ .  $\nu$  is the shape parameter estimated when errors are assumed to be non-normal.  $\phi$  is the persistence parameter given by  $(\alpha + 0.5\gamma + \beta)$  and SBIC is the Schwarz Bayesian Information Criterion which is calculated as  $-2 \times \ln L + k \times \ln(n)$ , where  $\ln L$  is the log-likelihood,  $k$  the number of parameters and  $n$  the number of observations. Robust standard errors are given in parentheses*



**Table 7: Parameter estimates for the FIGARCH (1,d,1) model**

Parameter	Distribution		
	Normal	Student's-t	GED
$\hat{\omega}$	0.013050e-4 (0.014321e-4)	0.065194e-4 (0.051979e-4)	0.049579e-4 (0.029973)
$\hat{\phi}$	-0.179579 (0.11596)	-0.121080 (0.23481)	0.000448 (0.14253)
$\hat{d}$	1.121334 (0.19068)	0.321404 (0.091072)	0.498707 (0.27242)
$\hat{\beta}$	0.916047 (0.077652)	0.168135 (0.29760)	0.469504 (0.38055)
$\hat{\nu}$		9.899157 (3.3205)	1.521196 (0.14762)
SBIC	-7887.56	-7944.43	-7911.59

*Note: The table presents the MLE for a FIGARCH (1,d,1) for the conditional variance for three possible error distributions. The conditional variance for the FIGARCH model is given by the ARCH( $\infty$ ) representation:  $h_t = \omega + \sum_{i=1}^{\infty} \lambda_i e_{t-i}^2$ , where  $\delta_1 = d, \lambda_1 = \phi - \beta + d, \delta_i = \frac{i-1-d}{i} \delta_{i-1}$  and  $\lambda_i = \beta \lambda_{i-1} + \delta_i - \phi \delta_{i-1}$  for  $i = 2, \dots$ . Note that  $\phi$  is now the AR coefficient. The degree of fractional integration is given by the parameter  $d$ . Robust Standard errors are given in parentheses.*

**Table 8: Parameter Estimates for HEAVY (1,1) model**

Parameter	
$\hat{\omega}$	5.5398e-6 (2.0289e-06)
$\hat{\alpha}$	0.4806 (0.0690)
$\hat{\beta}$	0.4908 (0.0721)
$\hat{\phi}$	0.9714

Note: The table presents the MLE for the HEAVY(1,1) model where the conditional variance is driven by the following process:  $h_t = \omega_H + \alpha_H(RV_{t-1}) + \beta_H(h_{t-1})$ , where  $RV_{t-1}$  is any given realized volatility measure. Note that  $\hat{\phi}$  is the persistence parameter ( $\hat{\alpha} + \hat{\beta}$ ).

**Table 9: Parameter Estimates for ARFIMA(2,d,2) model of the log (RV)**

Parameter	
$\hat{\mu}$	-9.12441
$\hat{\phi}_1$	0.619902 (0.01726)
$\hat{\phi}_2$	-0.98392 (0.01473)
$\hat{d}$	0.49093 (0.01078)
$\hat{\theta}_1$	-0.61572 (0.02835)
$\hat{\theta}_2$	0.959143 (0.02569)

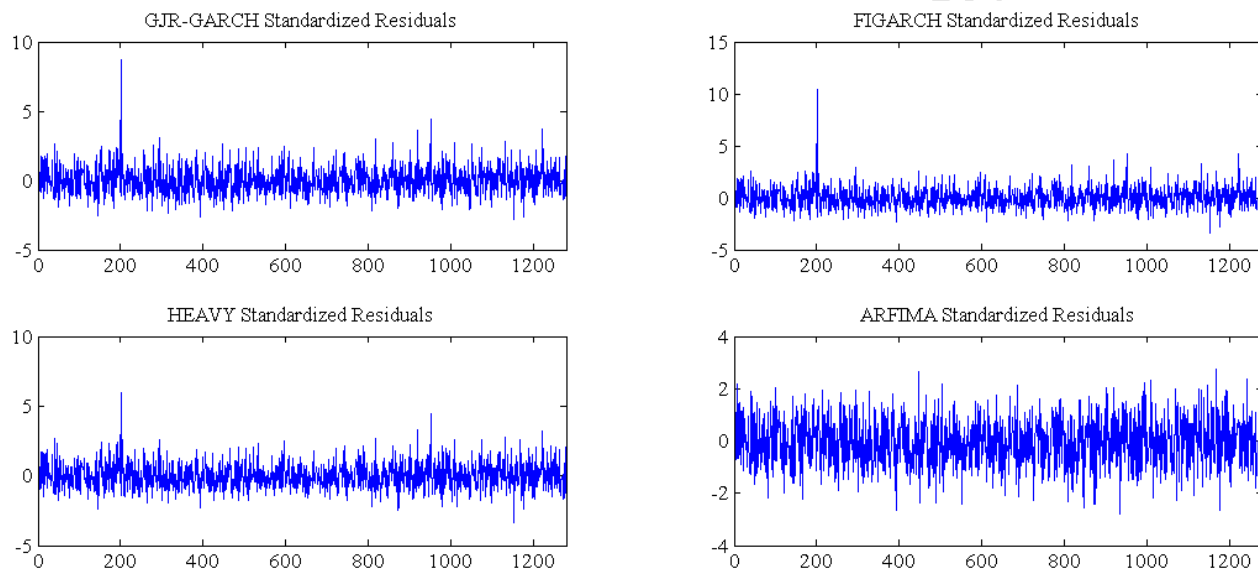
Note: This table presents the MLE for the ARFIMA (2,d,2) model of the log(RV) which is given by the following expression  $(1 - L)^d \phi(L) \log(RV_t) = \theta(L)e_t$  where  $\phi(L)$  and  $\theta(L)$  are lag polynomials of order 2. Robust standard errors are given in parentheses.

**Table 10: Diagnostic Information for Standardized Residuals**

	Model			
	GARCH	FIGARCH	HEAVY	ARFIMA-RV
Mean	0.0190	0.0214	0.0187	-0.0443
Standard Deviation	1.0128	1.0186	0.9467	0.9257
$LM(20)$	6.4483	5.2714	13.2022	15.2255
$Q^2(20)$	13.7232	9.5569	22.8945	28.7705

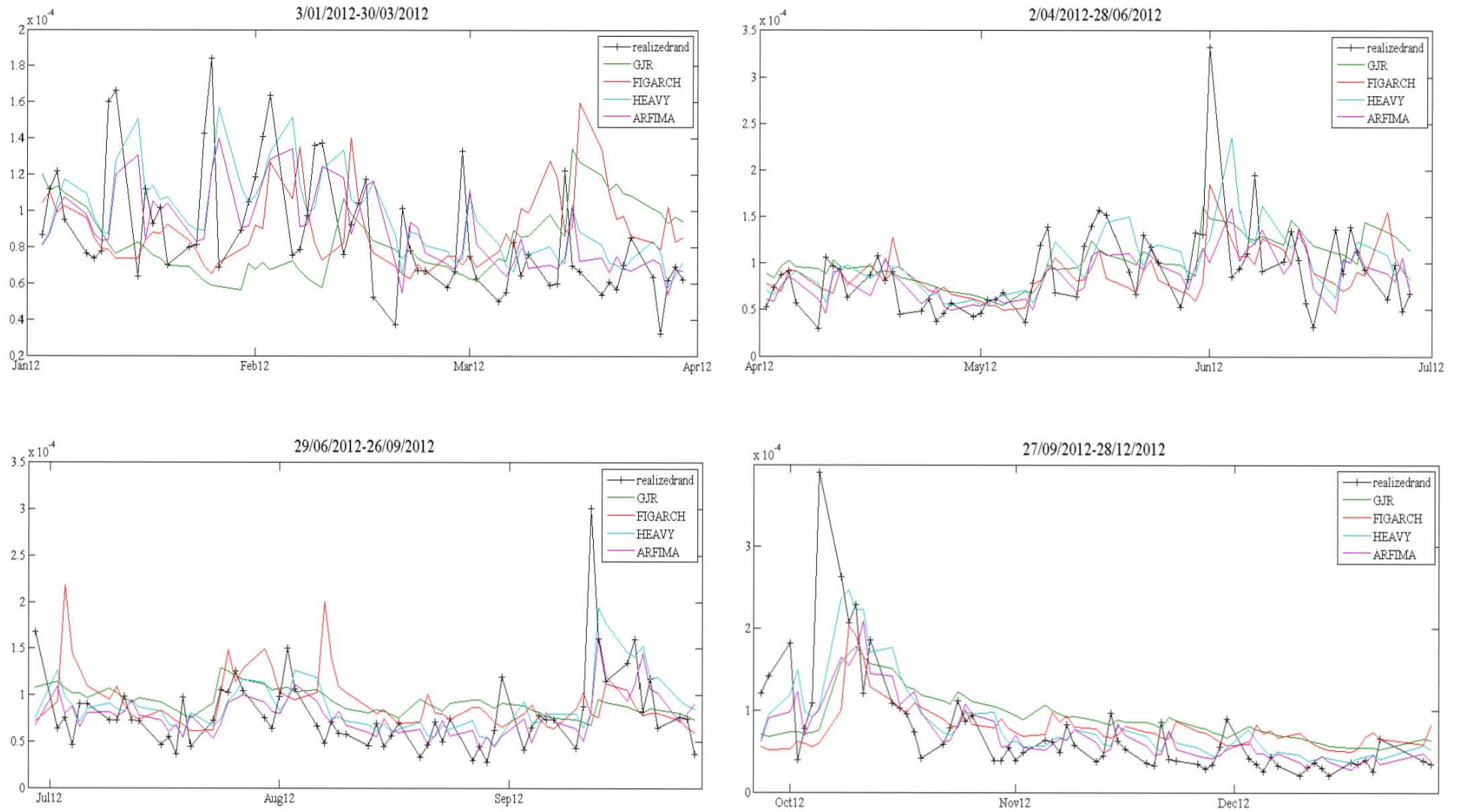
*Note: The table presents diagnostic information for the standardized residuals from the GARCH, FIGARCH, HEAVY and ARFIMA-RV models.  $LM(20)$  is the Lagrange Multiplier test statistic for ARCH effects up to the 20th lag which under the null hypothesis of no ARCH effects is  $\chi^2(20)$  distributed.  $Q^2(20)$  is the Ljung-Box  $Q$ -statistic that tests for serial autocorrelation in the squared residuals up to the 20th lag. Under the null of no serial autocorrelation this test is also  $\chi^2(20)$  distributed, thus we can accept the null for both of these tests for all the models at the 5% significance level since neither of them exceeds the critical value of 31.41.*

**Figure 10: Standardized Residuals**



*Note: The figure graphs the in-sample standardized residuals,  $z_t = \frac{r_t - \mu}{h_t^{1/2}}$ , from the GJR-GARCH, FIGARCH, HEAVY and ARFIMA-RV models.*

**Figure 11: Univariate Forecasts**



*Note: The figure graphs the 1-step ahead USDZAR daily conditional variances from the GJR-GARCH, FIGARCH, HEAVY and ARFIMA-RV models, as well as the MA(1)-based daily realized USDZAR variance. The period spans all the trading days in 2012.*

**Table 11: Mincer-Zarnowitz Regressions (Realized Variance)**

Explanatory Variables	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$R^2$
GJR-GARCH	0.0000240 (0.0000152)	0.6450522 (0.1623124)	- -	0.0987657
FIGARCH	0.0000435 (0.0000152)	0.4565371 (0.1742296)	- -	0.0630227
HEAVY	0.0000200 0.0000073	0.7009844 0.0866066	- -	0.2597202
ARFIMA-RV	0.0000098 (0.0000079)	0.9051450 (0.1081969)	- -	0.2787478
ARFIMA-RV & GJR-GARCH	0.0000029 (0.0000134)	0.8537334 (0.1273611)	0.1197705 (0.1774018)	0.2812535
ARFIMA-RV & FIGARCH	0.0000065 (0.0000129)	0.8812388 (0.1259387)	0.0597401 (0.1767098)	0.2796325
ARFIMA-RV & HEAVY	0.0000106 (0.0000075)	0.6980740 (0.2859938)	0.1778747 (0.2493392)	0.2808824

Notes: The table presents the Mincer-Zarnowitz regression estimates from the following OLS regression:  $\hat{\sigma}_t^2 = \hat{\beta}_0 + \hat{\beta}_1(h_{t,ARFIMA-RV}) + \hat{\beta}_2(h_{t,GARCH Model}) + e_t$ , where  $\hat{\sigma}_t^2$  is the MA(1) realized variance and  $h_t$  the conditional variance. 'GARCH model' represents one of the competing GARCH-type models; GJR-GARCH, FIGARCH, or HEAVY. The forecast evaluation period is 2012 or 256 trading days. Robust standard errors are given in parentheses.

**Table 12: Mincer-Zarnowitz Regressions (Daily Returns Squared)**

Explanatory Variables	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$R^2$
GJR-GARCH	0.0000757 (0.0000276)	0.0201045 (0.2792716)	- -	0.0000134
FIGARCH	0.0000987 (0.0000294)	-0.2398744 (0.2887329)	- -	0.0024241
HEAVY	0.0000483 (0.0000167)	0.3226315 (0.1799454)	- -	0.0076655
ARFIMA-RV	0.0000391 (0.0000188)	0.4719588 (0.2460748)	- -	0.0105590
ARFIMA-RV & GJR-GARCH	0.0000602 (0.0000247)	0.6296153 (0.3514078)	-0.3672826 (0.4041141)	0.0138420
ARFIMA-RV & FIGARCH	0.0000696 (0.0000273)	0.6927777 (0.3041936)	-0.5518127 (0.3524932)	0.0210758
ARFIMA-RV & HEAVY	0.0000380 (0.0000192)	0.7550276 (0.8422110)	-0.2431571 (0.6528312)	0.0111147

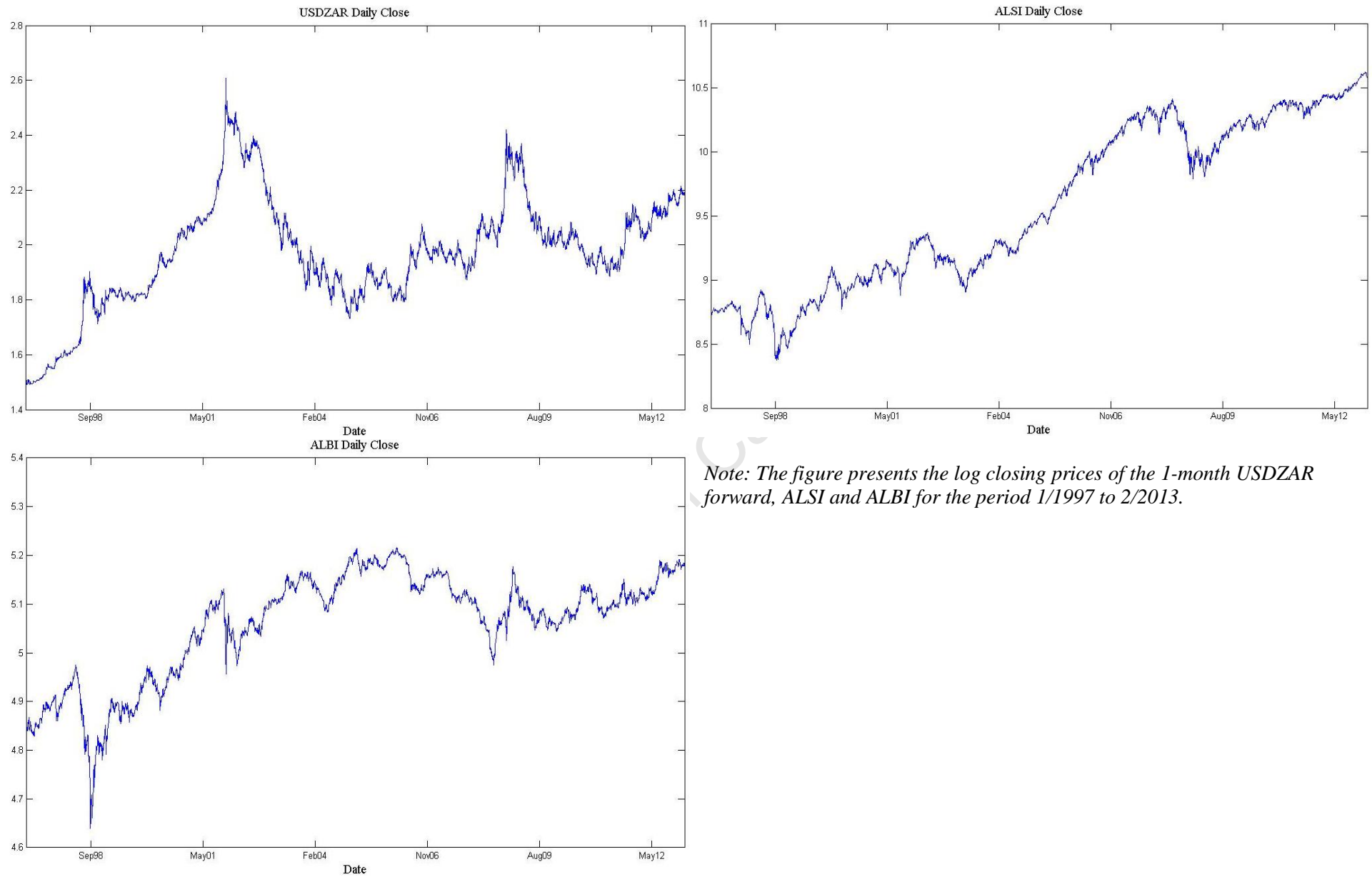
Notes: See Table 11 notes.  $\hat{\sigma}_t^2$  is now the daily return squared.

**Table 13: SPA test based on MSE and QLIKE**

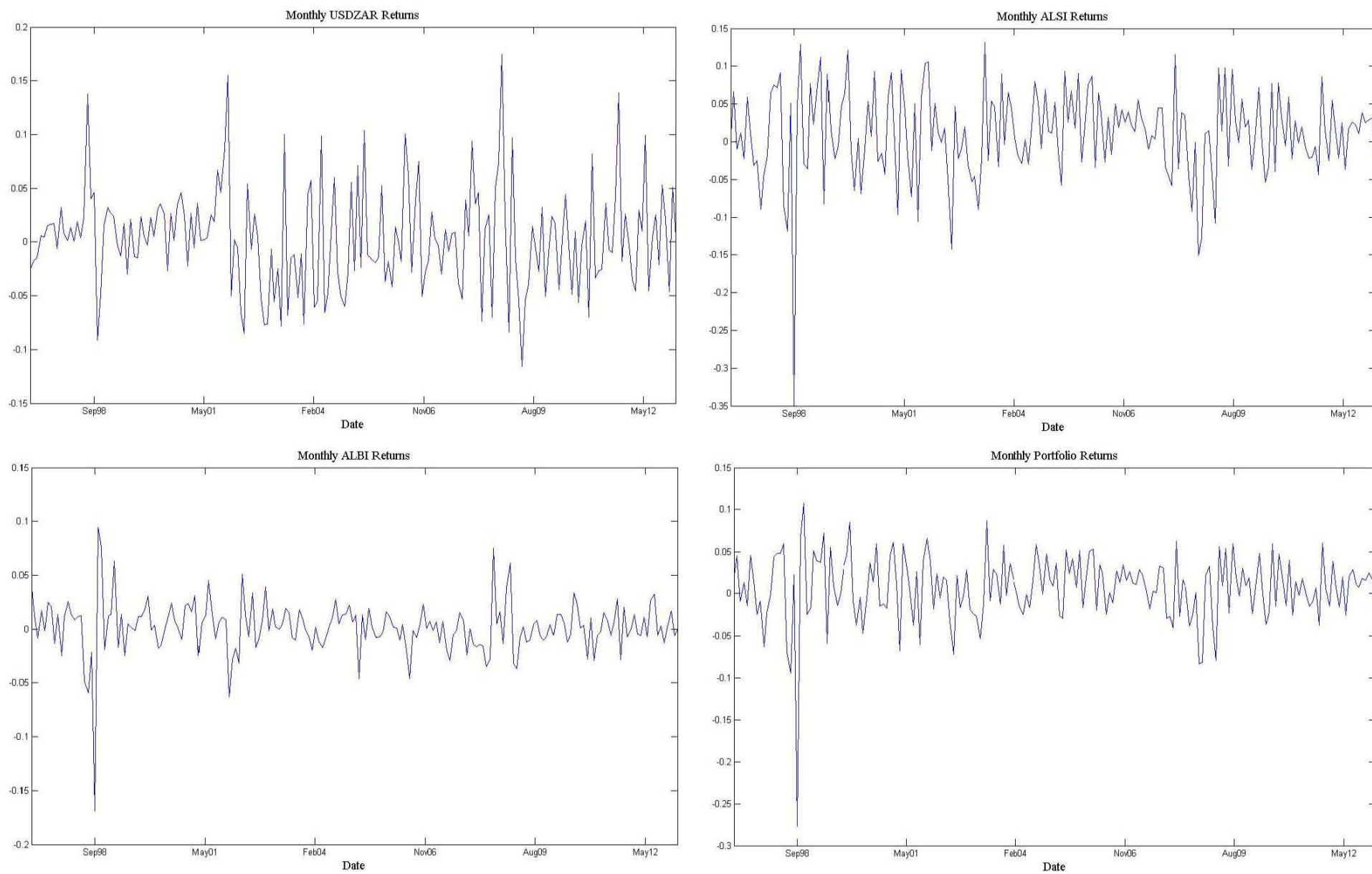
Benchmark	MSE			QLIKE		
	Consistent p-val	Upper p-val	Lower p-val	Consistent p-val	Upper p-val	Lower p-val
GJR-GARCH	0	0	0	0.0012	0.0012	0.0006
FIGARCH	0.0002	0.0002	0.0002	0.0044	0.0044	0.0044
HEAVY	0.0024	0.0024	0.0024	0.0082	0.0198	0.0082
ARFIMA	1	1	1	1	1	1

*Note: The table presents the Superior Predictive Ability (SPA) test's Consistent, Upper and Lower p-values. The null hypothesis is that no alternative model outperforms the benchmark. The number of bootstrap replications is 5000 and the average block length was chosen based on the procedure described in Patton, Politis and White, (2004). For the MSE the block length was 1 and for the QLIKE, 20.*

**Figure 12: Daily log closing prices**



**Figure 13: Monthly Returns**



*Note: The figure presents the monthly USDZAR forward, ALSI, ALBI and Mixed Portfolio returns.*



**Table 14: Descriptive Statistics for Returns**

Daily Returns										
	Mean(%)	Standard Deviation	Skewness	Kurtosis	Min	Max	<i>JB</i>	<i>AC</i> (1)	<i>LM</i> (20)	$Q^2$ (20)
ALBI	9.0362e-05	0.0049	-1.1141	27.4093	-0.0785	0.0508	105424*	0.1012*	718*	1580*
ALSI	4.5073e-04	0.0129	-0.4711	9.1530	-0.1269	0.0742	6797*	0.0676*	631*	1872*
Mixed Portfolio	3.0658e-04	0.0083	-0.5730	9.8620	-0.0827	0.0501	8490*	0.0703*	726*	2081*
Monthly Returns										
USDZAR	0.0033	0.0474	0.5631	3.9836	-0.1161	0.1747	17.19*	0.0421	18.2279	22.149
ALSI	0.0097	0.0599	-1.2377	8.6978	-0.3489	0.1313	301*	-0.0209	11.77	15.5949
ALBI	0.0020	0.0249	-1.2282	14.7653	-0.1690	0.0947	1132*	0.023	22.2568	32.4397*
Mixed Portfolio	0.0066	0.0406	-1.8432	14.0828	-0.2769	0.1079	1069*	-0.031	10.9008	13.6472

Notes: See Table 1 notes. The daily returns were calculated as the difference in the log closing prices. The sample period begins on the 31/12/1996 and ends on the 1/03/2013 for a total of 4218 observations.

**Table 15: Correlation, 1/1997-2/2013**

	USDZAR	Portfolio	ALSI	ALBI
USDZAR	1			
Mixed Portfolio	-0.26901	1		
ALSI	-0.19242	0.973401	1	
ALBI	-0.40206	0.565213	0.361176	1

Note: The table presents the correlation for the period January 1997-February 2013.

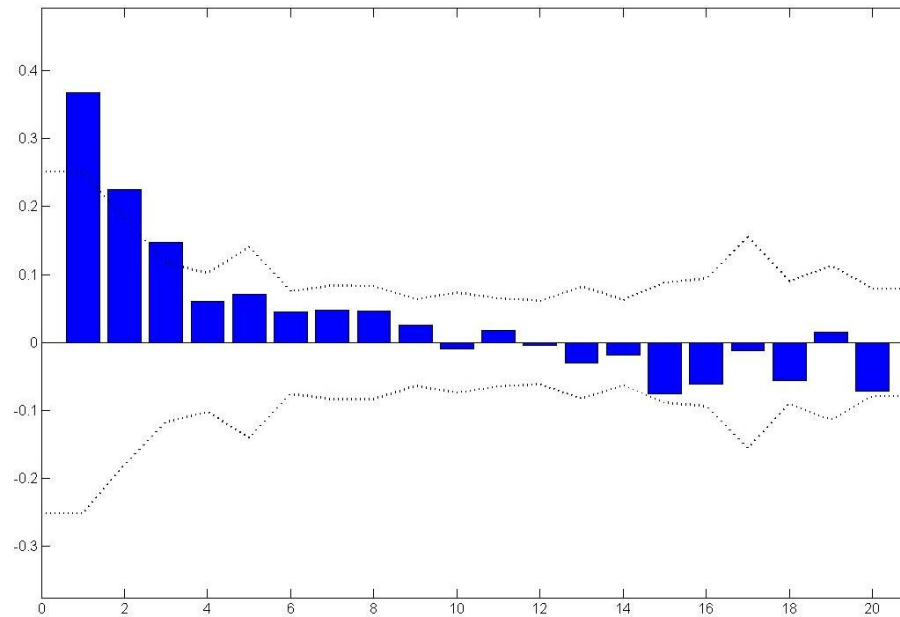
**Table 16: Correlation, 1/2008-12/2009**

	USDZAR	Portfolio	ALSI	ALBI
USDZAR	1			
Mixed Portfolio	-0.49205	1		
ALSI	-0.41433	0.968461	1	
ALBI	-0.32787	0.166103	-0.08484	1

Note: The table presents the correlation for the period January 2008-December 2009.

**Table 17: Descriptive Statistics for Monthly Realized Volatility and log(RV)**

	Mean	Standard Deviation	Skewness	Kurtosis	Min	Max	<i>JB</i>	<i>AC</i> (1)	<i>Q</i> (20)
USDZAR									
RV	0.0023	0.0031	5.5843	44.6850	4.0252e-5	0.0299	14694*	0.3691*	49*
Log(RV)	-6.5829	1.1008	-0.7325	4.1328	-10.1203	-3.5101	26.41*	0.7283*	403*
ALSI									
RV	0.0036	0.0045	3.8067	21.4373	0.000265	0.0340	3111*	0.4874*	46*
Log(RV)	-6.0554	0.8701	0.4577	3.3100	-8.2351	-3.3806	7.2097*	0.6028*	190*
Portfolio									
RV	0.0015	0.0020	3.8541	21.0358	1.4009e-04	0.0144	3008*	0.4689*	93*
Log(RV)	-6.9496	0.8528	0.7273	3.5175	-8.8732	-4.2422	18.52*	0.6061*	203*

**Figure 14: Sample ACF, Realized USDZAR Variance (monthly)**

*Note: The figure presents the sample ACF for monthly USDZAR 1-month forward realized variance.*

**Table 18: Descriptive Statistics for Standardized Returns (monthly)**

	Mean	Standard Deviation	Skewness	Kurtosis	Min	Max	<i>JB</i>	<i>AC</i> (1)	Q(20)
USDZAR	-0.0815	1.0197	0.1342	2.3867	-2.4228	2.2835	3.82	0.0273	12.37
ALSI	0.1055	1.0331	0.0741	2.3326	-2.6804	2.3572	3.9971	-0.0431	12.6402
Mixed Portfolio	0.1245	1.0543	0.0417	2.3264	-2.7594	2.3965	3.9475	-0.0116	12.8532

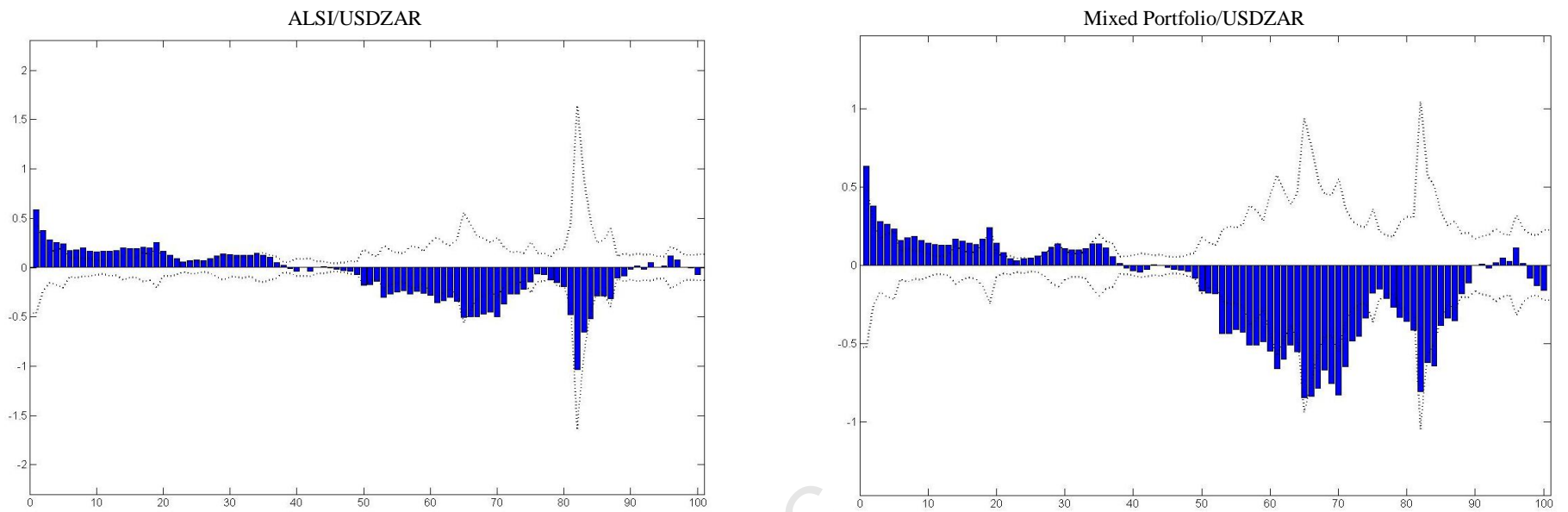
*Note: See Table 1 notes.*

**Table 19: Descriptive Statistics for Realized Covariance**

$w_{ALSI}, w_{ALBI}$	Mean	Standard Deviation	Skewness	Kurtosis	Min	Max	<i>JB</i>	<i>AC</i> (1)	Q(20)
1, 0	-5.2309e-04	0.0023	-2.8431	34.3727	-0.0197	0.0123	7951*	0.5874*	248*
0.6, 0.4	-5.0429e-04	0.0014	-5.1308	45.6258	-0.0138	0.0024	15046*	0.6348*	229*

*Note: See Table 1 notes.*

**Figure 15: Realized Covariances**



*Note: The figure presents the sample ACF for the monthly realized covariance between the 1-month USDZAR forward and ALSI or the Mixed Portfolio.*

**Table 20: Scalar BEKK parameter estimates**

Parameter	ALSI	Mixed Portfolio
$\hat{C}_{11}$	0.01354000 (0.0041304)	0.01302800 (0.0042828)
$\hat{C}_{12}$	-0.00523900 (0.0029119)	-0.00381400 (0.0020619)
$\hat{C}_{22}$	0.01493900 (0.0050823)	0.00930500 (0.0035073)
$\hat{a}$	0.43488200 (0.10785)	0.40057000 (0.10467)
$\hat{b}$	0.87358800 (0.058992)	0.88874900 (0.057004)
$\hat{\nu}$	7.27887200 (2.3733)	6.66771400 (2.2188)
SBIC	-1204.91000000	-1378.25

*Note: The table presents the MLE for the Scalar BEKK model where conditional covariance process for the bivariate returns are given by  $H_t = C'C + a^2 e_{t-1} e'_{t-1} + b^2 H_{t-1}$ , where  $C$  is a lower triangular matrix.  $\hat{\nu}$  is the shape parameter. Robust standard errors are given in parentheses.*

**Table 21: DCC parameter estimates**

Parameter	ALSI			Portfolio		
	USDZAR	ALSI	Correlation	USDZAR	Portfolio	Correlation
$\hat{\omega}$	0.00041333 (0.0004649)	0.00024535 (0.0001479)		0.00041333 (0.0004649)	0.00011447 (0.0000881)	
$\hat{\alpha}$	0.20031216 (0.2209606)	0.07718349 (0.0746429)	0.07314180 (0.0518369)	0.20031216 (0.2209606)	0.09409424 (0.1071186)	0.06913805 (0.0337282)
$\hat{\gamma}$		0.53444073 (0.1839605)			0.38596575 (0.1542282)	
$\hat{\beta}$	0.62968390 (0.3518646)	0.64726611 (0.0774533)	0.88295224 (0.0810307)	0.62968390 (0.3518646)	0.67132160 (0.0974178)	0.90076913 (0.0432946)
$\hat{\phi}$	0.82999606	0.99166997		0.82999606	0.95839872	
$\hat{A}$		7.9242882			5.1019062	
$\hat{\rho}$			-0.24015716 (0.0793675)			-0.31444853 (0.084406)
SBIC	-1196.26687731			-1366.67123		

Note: The table presents the MLE for the DCC model. The conditional variance for the USDZAR forward is given by:  $h_t = \omega + \alpha e_{t-1}^2 + \beta h_{t-1}$ , the conditional variance for the ALSI/Portfolio by:  $h_t = \omega + \alpha(e_{t-1}^2) + I_{e_{t-1} < 0} \gamma(e_{t-1}^2) + \beta(h_{t-1})$ , where  $I_{e_{t-1} < 0}$  is an indicator function. The conditional correlation is given by  $\rho_{i,j,t} = \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}}\sqrt{q_{j,j,t}}}$  where the process for the conditional variance of the standardized residuals  $z_t$  is given by  $q_{i,j,t} = \rho_{i,j}(1 - \alpha - \beta) + \alpha z_{i,t-1} z_{j,t-1} + \beta q_{i,j,t-1}$ .  $\hat{A}$  is the asymmetry ratio ( $\frac{\hat{\alpha} + \hat{\gamma}}{\hat{\alpha}}$ ). Robust standard errors are given in parentheses.

**Table 22: GO-GARCH parameter estimates**

Parameter	ALSI		Portfolio	
	Factor 1	Factor 2	Factor 1	Factor 2
$\hat{\alpha}$	0.3285390 (0.0769450)	0.2609690 (0.0780510)	0.3278820 (0.0688340)	0.2599760 (0.0617890)
$\hat{\beta}$	0.5301360 (0.0962490)	0.4783350 (0.1357000)	0.5208410 (0.0927390)	0.5573190 (0.1060100)
SBIC	-1232.773		-1380.3294000	

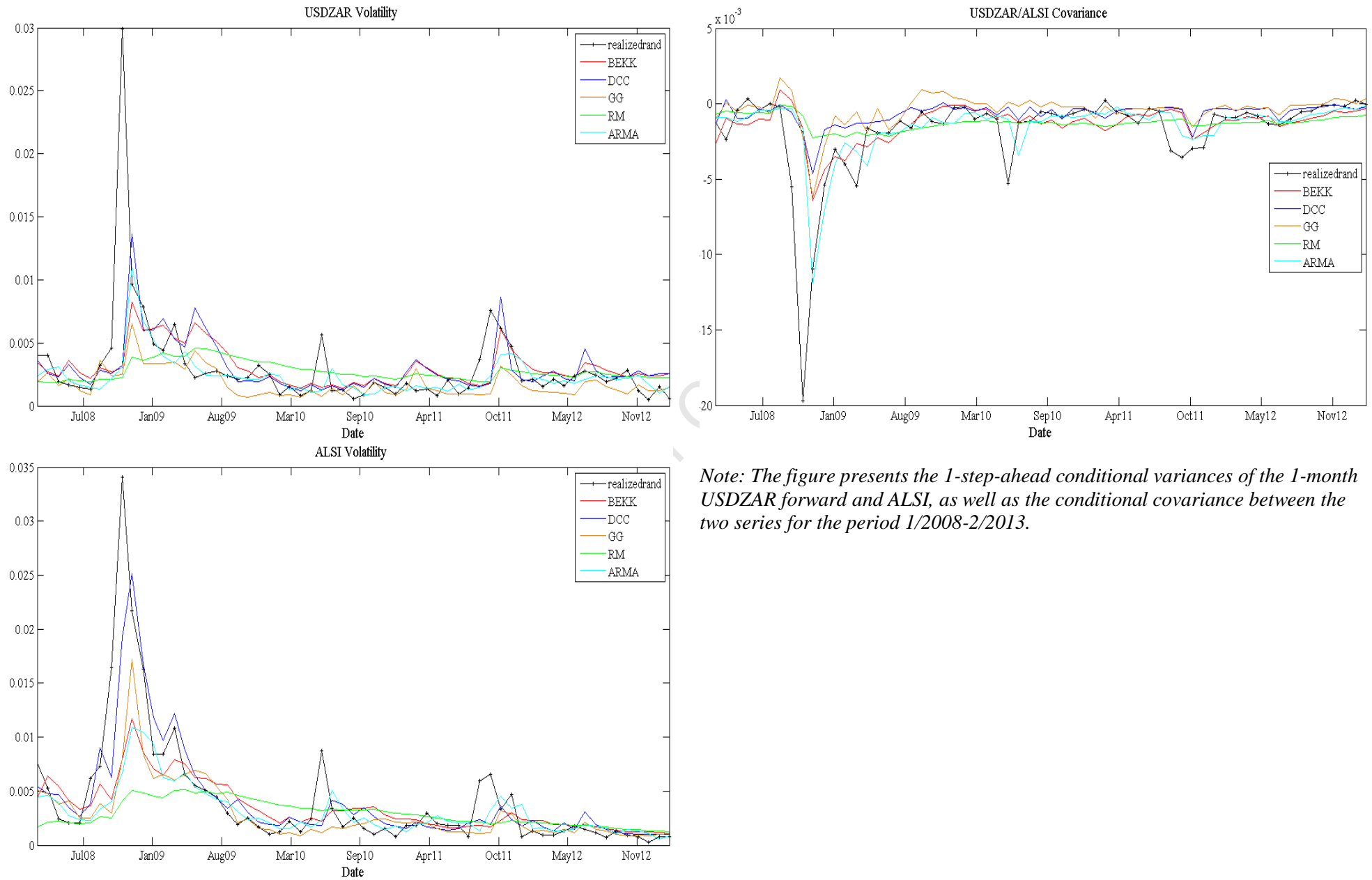
Note: The table presents the Non-Linear Least Squares estimates for the GO-GARCH parameters. The conditional covariance process is given by  $H_t = WH_t^f W'$ , where  $H_t^f$  is the conditional variance for the unobserved factor (vector) process  $f$  and is given by  $H_t^f = (I - A - B) + A \odot (f_{t-1} f'_{t-1}) + BH_{t-1}^f$

**Table 23: ARFIMA (p,d,q) estimates for Realized Covariance Matrix**

Parameter	Log(realizedrand)	Log(realizedalsi)	Log(realizedportfolio)	covrandalsi	covrandportfolio
$\hat{\mu}$	-6.6354700 (0.2486000)	-6.0979000 (0.1635000)	-6.9836600 (0.1571000)	-0.0005324 (0.0006708)	-0.0005025 (0.0005241)
$\hat{\phi}_1$	0.8449450 (0.0501600)	0.7961680 (0.0693100)	0.7889210 (0.0712500)	0.2886480 (0.1356000)	
$\hat{d}$				0.2427460 (0.1068000)	0.2960700 (0.0704900)
$\hat{\theta}_1$	-0.2510680 (0.0891900)	-0.3042380 (0.1056000)	-0.2943010 (0.1102000)		0.3330570 (0.0823100)

Notes: This table presents the MLE for the ARFIMA (1,d,1) model of the unique elements of the covariance matrix which is given by the following expression  $(1 - L)^d \phi(L)x_t = \theta(L)e_t$ , where  $\phi(L)$  and  $\theta(L)$  are lag polynomials of order 1. Robust Standard Errors are given in parentheses.

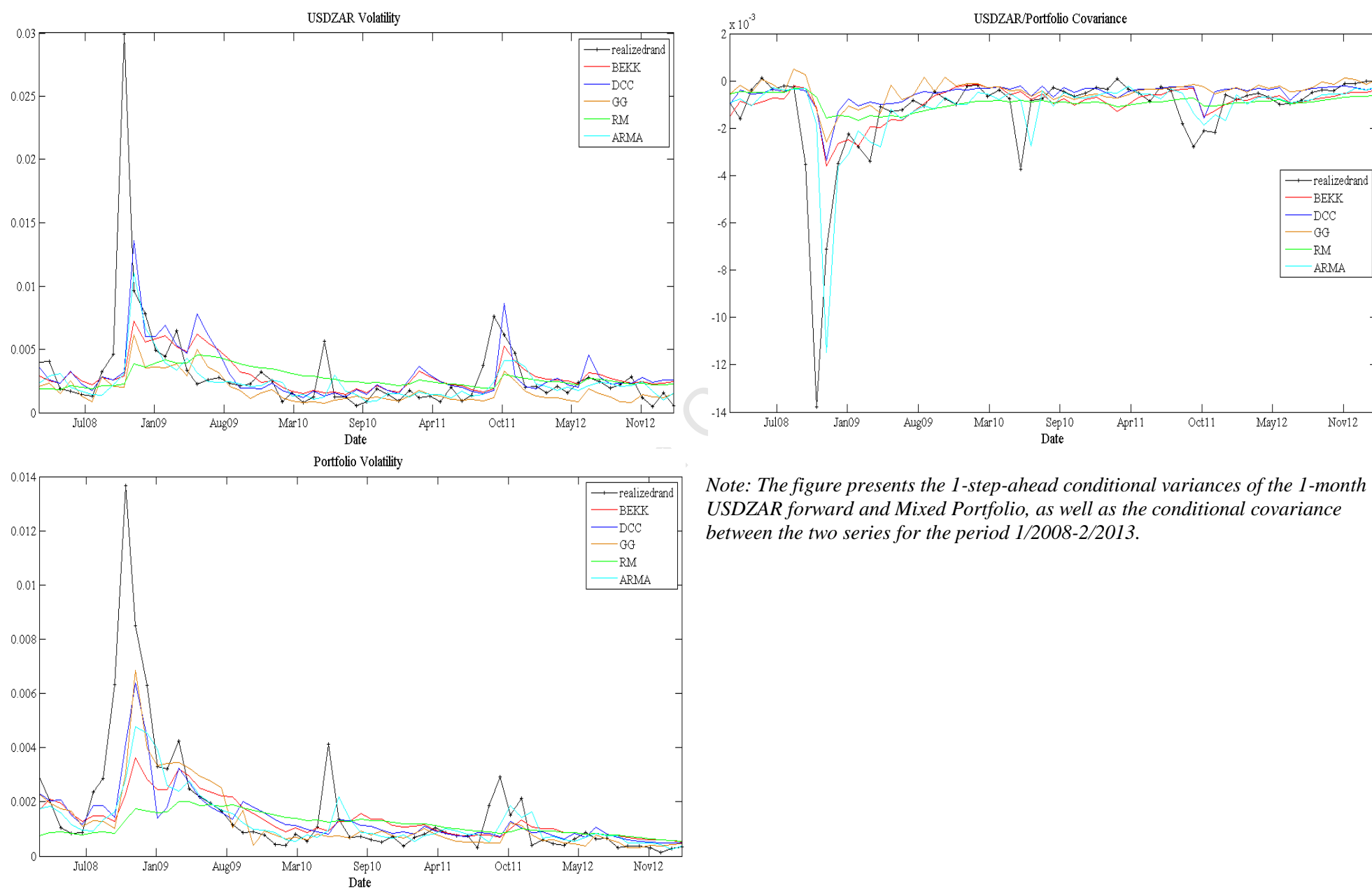
**Figure 16: USDZAR/ALSI Covariance Matrix**



*Note: The figure presents the 1-step-ahead conditional variances of the 1-month USDZAR forward and ALSI, as well as the conditional covariance between the two series for the period 1/2008-2/2013.*

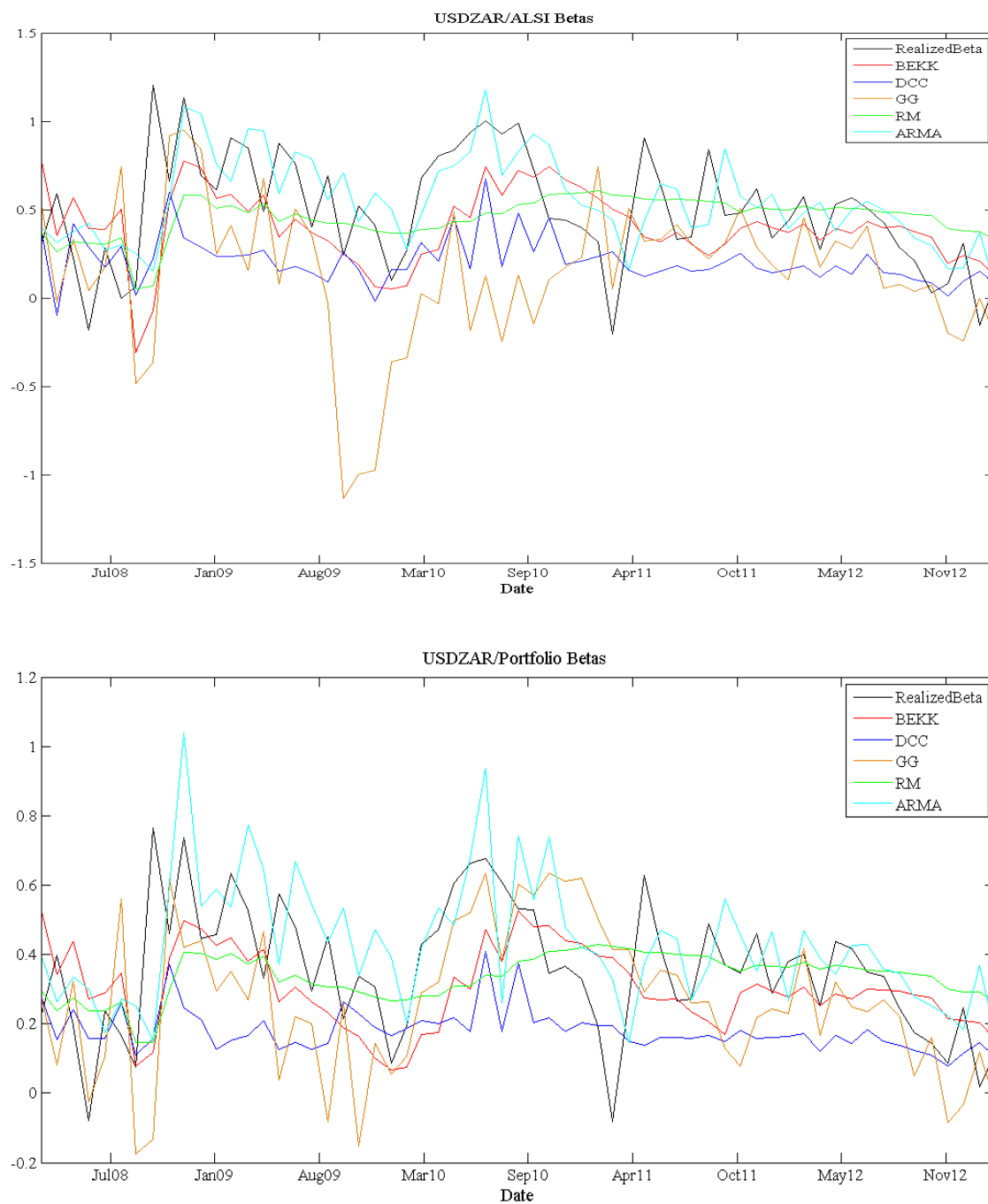


**Figure 17: USDZAR/Portfolio Covariance Matrix**



*Note: The figure presents the 1-step-ahead conditional variances of the 1-month USDZAR forward and Mixed Portfolio, as well as the conditional covariance between the two series for the period 1/2008-2/2013.*

**Figure 18: Betas**



*Note: The figure presents the conditional betas from the BEKK, DCC, GO-GARCH, RiskMetrics and ARFIMA-RV model for the period 1/2008-2/2013.*

**Table 24: SPA test based on Euclidean distance, Full sample**

Benchmark	ALSI			Portfolio		
	Consistent p-val	Upper p-val	Lower p-val	Consistent p-val	Upper p-val	Lower p-val
BEKK	0.2656	0.3346	0.2032	0.0468	0.0476	0.0414
DCC	1	1	1	0.0562	0.0562	0.0482
GO-GARCH	0.2150	0.2150	0.1768	0.0502	0.0502	0.0480
RM	0.0304	0.0304	0.0304	0.0392	0.0392	0.0392
ARFIMA	0.2676	0.4024	0.1658	1	1	1

*Note: The table presents the Superior Predictive Ability (SPA) test's Consistent, Upper and Lower p-values. The null hypothesis is that no alternative model outperforms the benchmark. The number of bootstrap replications is 5000 and the average block length was chosen based on the procedure described in Patton, Politis and White, (2004). For both the ALSI and Mixed Portfolio the average block length was one. The Euclidean weighting matrix is the  $I_3$  matrix. See main text for details.*

**Table 25: SPA test based on Euclidean distance, Full sample less September, October, November 2008**

Benchmark	ALSI			Portfolio		
	Consistent p-val	Upper p-val	Lower p-val	Consistent p-val	Upper p-val	Lower p-val
BEKK	0.1018	0.1628	0.1018	0.0674	0.0674	0.0308
DCC	0.3634	0.3634	0.1896	0.0364	0.0364	0.0340
GO-GARCH	0.0194	0.0194	0.0168	0.0474	0.0474	0.0318
RM	0.0278	0.0278	0.0278	0.0200	0.0200	0.0200
ARFIMA	1	1	1	1	1	1

*Note: See Table 24 notes.*

**Table 26: SPA test based on Frobenius distance, Full sample**

Benchmark	ALSI			Portfolio		
	Consistent p-val	Upper p-val	Lower p-val	Consistent p-val	Upper p-val	Lower p-val
BEKK	0.2964	0.3898	0.2268	0.0634	0.0848	0.0634
DCC	1	1	1	0.0174	0.0174	0.0172
GO-GARCH	0.2054	0.2054	0.1736	0.0458	0.0458	0.0444
RM	0.0368	0.0368	0.0368	0.0522	0.0522	0.0522
ARFIMA	0.3164	0.4834	0.2022	1	1	1

*Note: See Table 24 notes.*

**Table 27: SPA test based on Frobenius distance, Full sample less September, October, November 2008**

Benchmark	ALSI			Portfolio		
	Consistent p-val	Upper p-val	Lower p-val	Consistent p-val	Upper p-val	Lower p-val
BEKK	0.0890	0.1374	0.0806	0.0542	0.0640	0.0256
DCC	0.1002	0.1252	0.0876	0.0238	0.0238	0.0232
GO-GARCH	0.0134	0.0134	0.0118	0.0452	0.0452	0.0344
RM	0.0300	0.0300	0.0300	0.0250	0.0250	0.0250
ARFIMA	1	1	1	1	1	1

*Note: See Table 24 notes.*

**Table 28: Ex-post utility, ALSI**

$\gamma$	Unhedged	BEKK	DCC	GO- GARCH	RiskMetrics	ARFIMA	OLS
1	0.000767	0.001741	0.002635	0.003215	0.00222	0.001949	0.00122
Rank	7	5	2	<u>1</u>	3	4	6
2	-0.00356	-0.00162	-0.00097	-0.00075	-0.00115	-0.00131	-0.00286
Rank	7	5	2	<u>1</u>	3	4	6
3	-0.00788	-0.00498	-0.00457	-0.00471	-0.00452	-0.00458	-0.00694
Rank	7	5	2	4	<u>1</u>	3	6
4	-0.01221	-0.00835	-0.00817	-0.00868	-0.00789	-0.00784	-0.01102
Rank	7	4	3	5	2	<u>1</u>	6
5	-0.01654	-0.01171	-0.01177	-0.01264	-0.01126	-0.01111	-0.0151
Rank	7	3	4	5	2	<u>1</u>	6
6	-0.02086	-0.01507	-0.01537	-0.0166	-0.01462	-0.01437	-0.01918
Rank	7	3	4	5	2	<u>1</u>	6

*Note: The table presents the rankings (based on average monthly ex-post utility) of the investor's unhedged portfolio or hedged portfolio, according to one of the conditional hedges or the conventional hedge.*

**Table 29: Annualized Variance, ALSI**

	Unhedged	BEKK	DCC	GO- GARCH	RiskMetrics	ARFIMA	OLS
Variance(%)	5.190716	4.035325	4.322064	4.75631	4.042615	3.916442	4.896299
Rank	7	2	4	5	3	<u>1</u>	6
Reduction(%)		28.63193	20.09808	9.133264	28.39998	32.53653	6.013062

*Note: The table presents the rankings of the unhedged and hedged portfolios based on their average monthly (annualized) variances. The percentage of variance reduction for the hedged portfolio relative to the unhedged portfolio is also given.*

**Table 30: Ex-post utility, Mixed Portfolio**

$\gamma$	Unhedged	BEKK	DCC	GO- GARCH	RiskMetrics	ARFIMA	OLS
1	0.001796	0.002778	0.003035	0.002187	0.003874	0.003905	0.003482
Rank	7	5	4	6	2	<u>1</u>	3
2	6.8E-05	0.001547	0.001699	0.00087	0.002641	0.002665	0.002141
Rank	7	5	4	6	2	<u>1</u>	3
3	-0.00166	0.000315	0.000362	-0.00045	0.001408	0.001425	0.0008
Rank	7	5	4	6	2	<u>1</u>	3
4	-0.00339	-0.00092	-0.00097	-0.00176	0.000175	0.000186	-0.00054
Rank	7	4	5	6	2	<u>1</u>	3
5	-0.00512	-0.00215	-0.00231	-0.00308	-0.00106	-0.00105	-0.00188
Rank	7	4	5	6	2	<u>1</u>	3
6	-0.00684	-0.00338	-0.00365	-0.0044	-0.00229	-0.00229	-0.00322
Rank	7	4	5	6	<u>1</u>	2	3

Note: See Table 28 notes.

**Table 31: Annualized Variance, Mixed Portfolio**

	Unhedged	BEKK	DCC	GO- GARCH	RiskMetrics	ARFIMA	OLS
Variance(%)	2.073292	1.478066	1.603750	1.580585	1.479605	1.487600	1.608890
Rank	7	<u>1</u>	5	4	2	3	6
Reduction(%)	-	40.270596	29.277732	31.172427	40.124718	39.371614	28.864715

Note: See Table 29 notes.